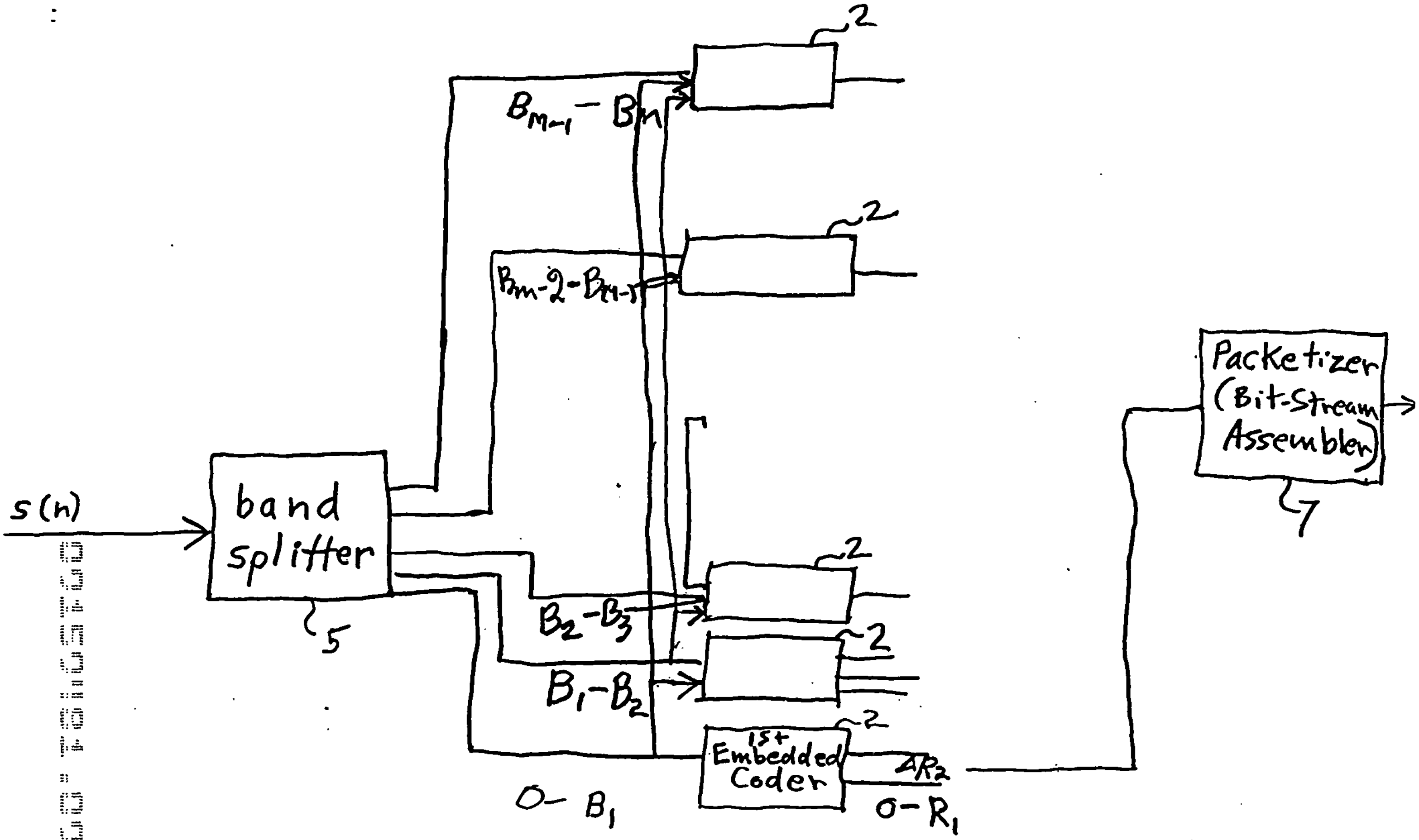

s(n)
band splitter
5
$B_{M-1}-B_M$
2
$B_{M-2}-B_{M-1}$
2
Packetizer (Bit-Stream Assembler)
7
B_2-B_3
2
B_1-B_2
2
1st Embedded Coder
2
ΔR_2
$0-B_1$
$0-R_1$

Fig. 1A

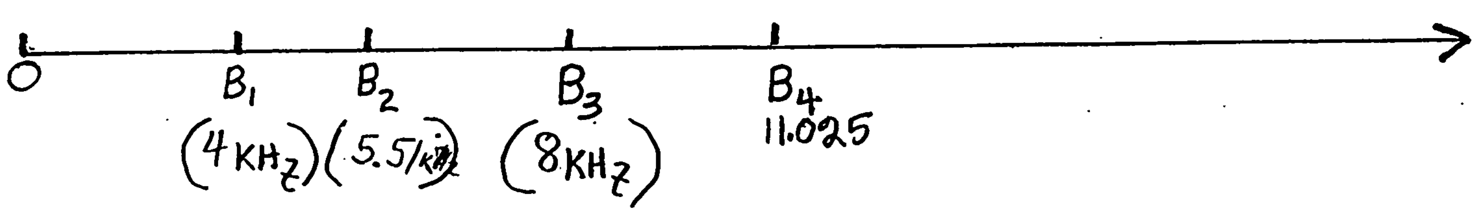

0
B_1
(4 KHz)
B_2
(5.5 KHz)
B_3
(8 KHz)
B_4
11.025

Fig. 1B

FFT-based Scalable and Embedded Codec Architecture
— Encoder with M Octave Bands

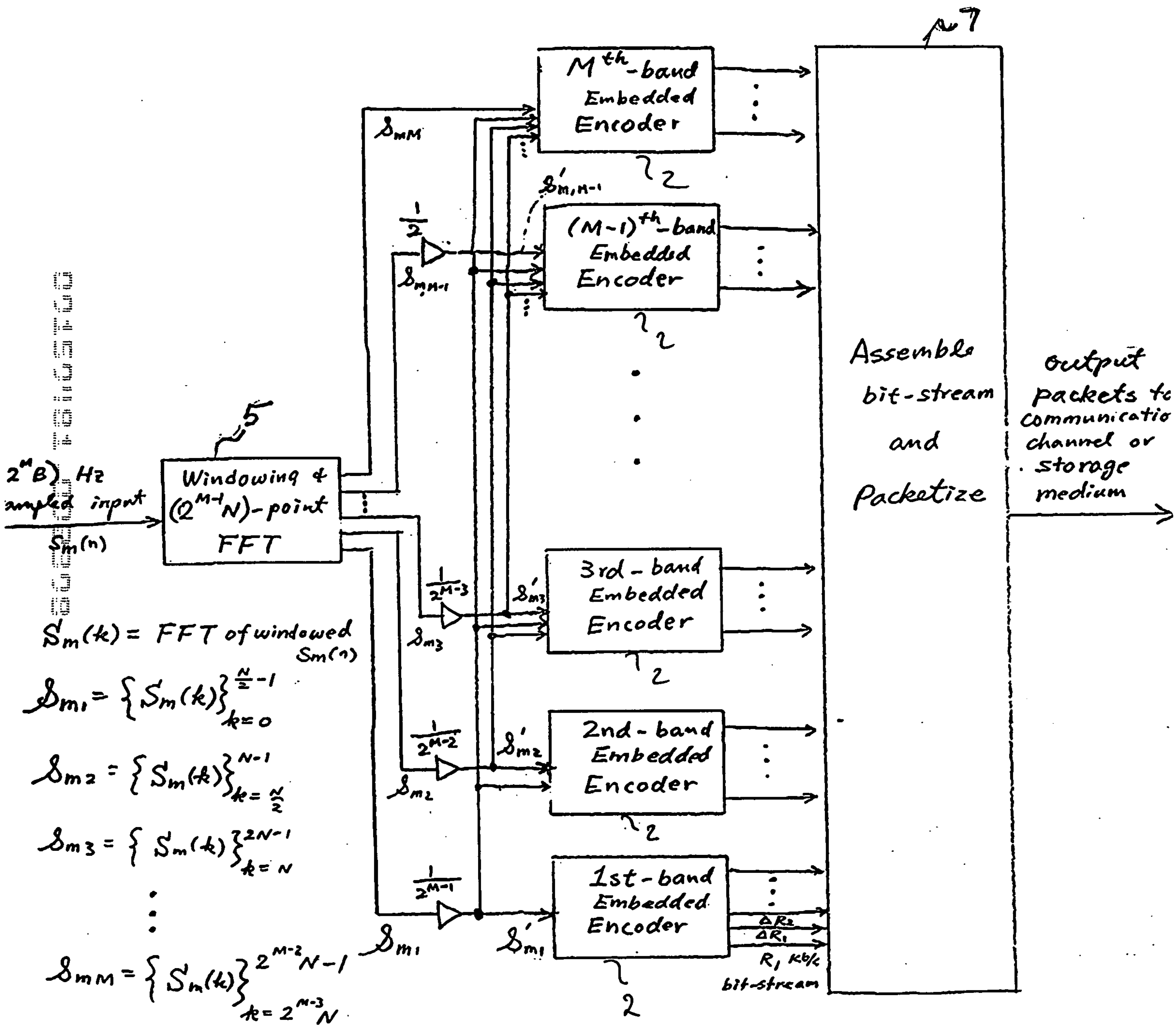


Fig. 2A

FFT-based Scalable and Embedded Codec Architecture
— Decoder with M_1 Octave Bands $(1 \le M_1 \le M)$

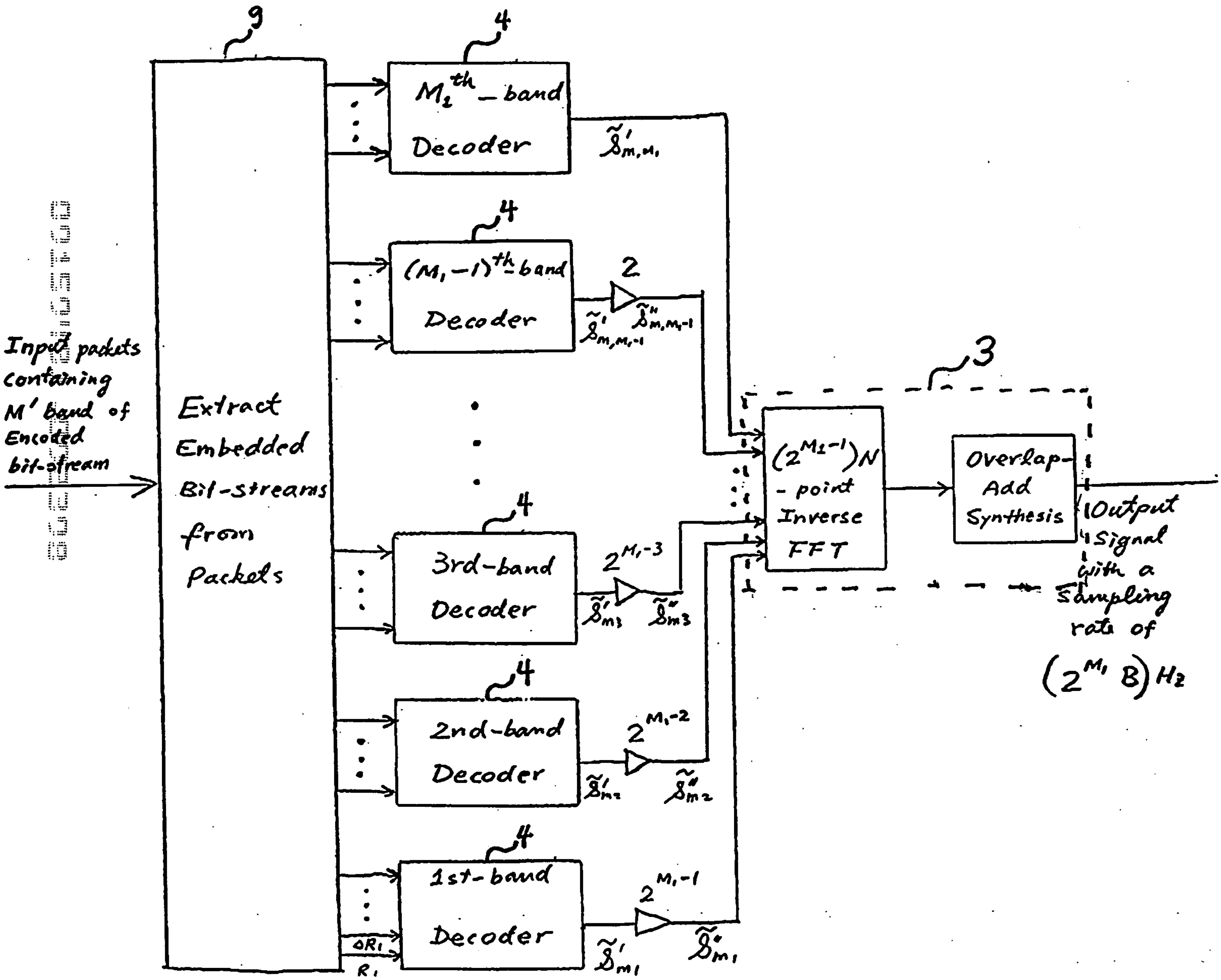


Fig. 2B

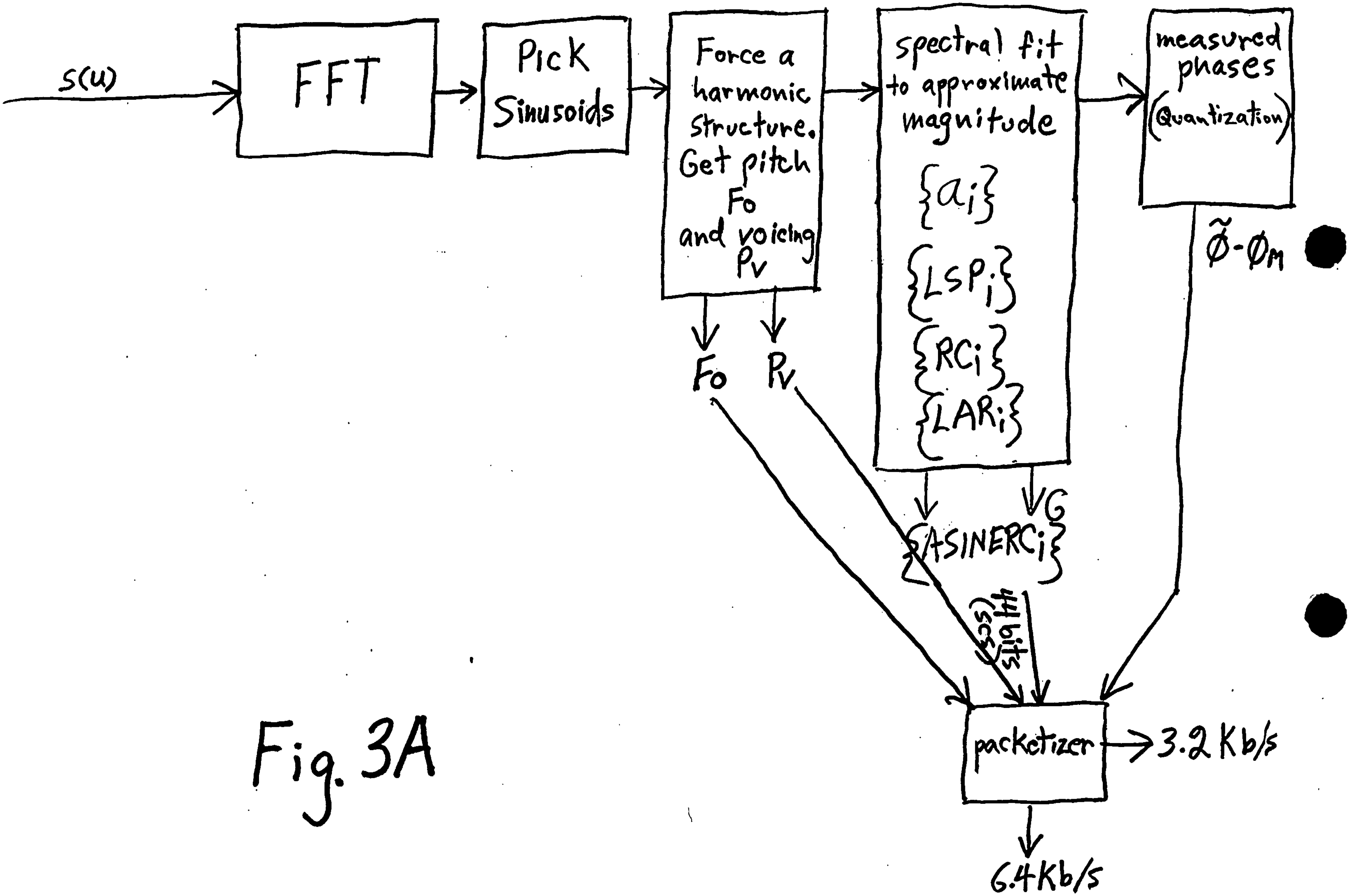
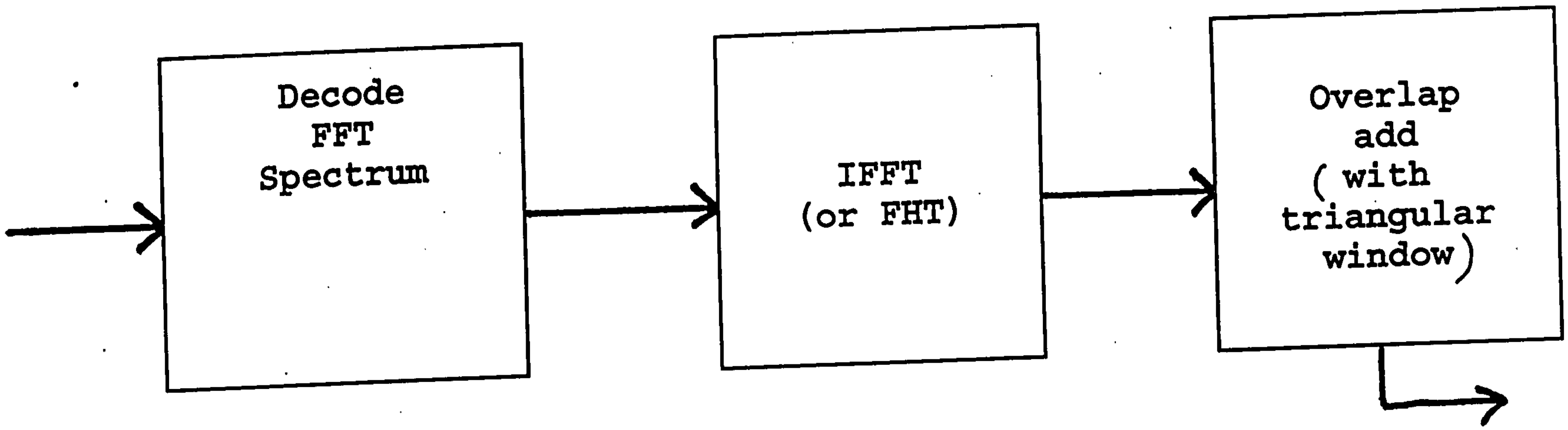

s(u)
FFT
Pick Sinusoids
Force a harmonic structure. Get pitch Fo and voicing Pv
Fo
Pv
spectral fit to approximate magnitude
{ai}
{LSPi}
{RCi}
{LARi}
{ASINERCi}
G
44 bits (scs)
measured phases (Quantization)
packetizer
3.2 Kb/s
6.4Kb/s

Fig. 3A



Decoder: Synthesis every M ms.

Fig. 3B

Bit Stream

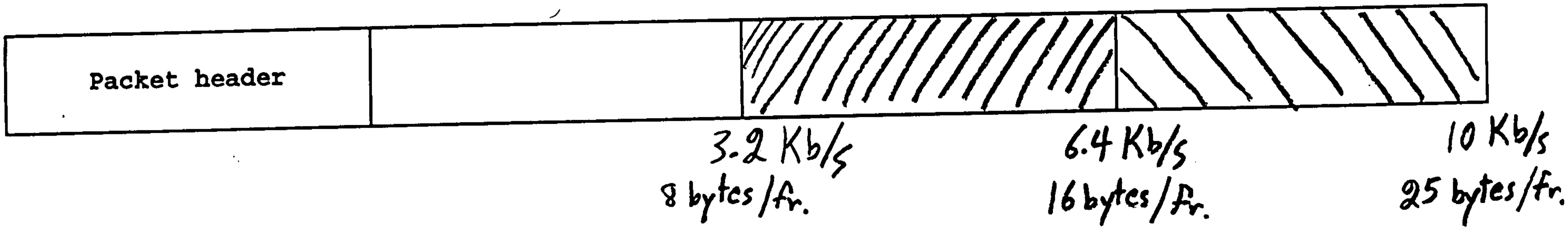


Fig. 4A

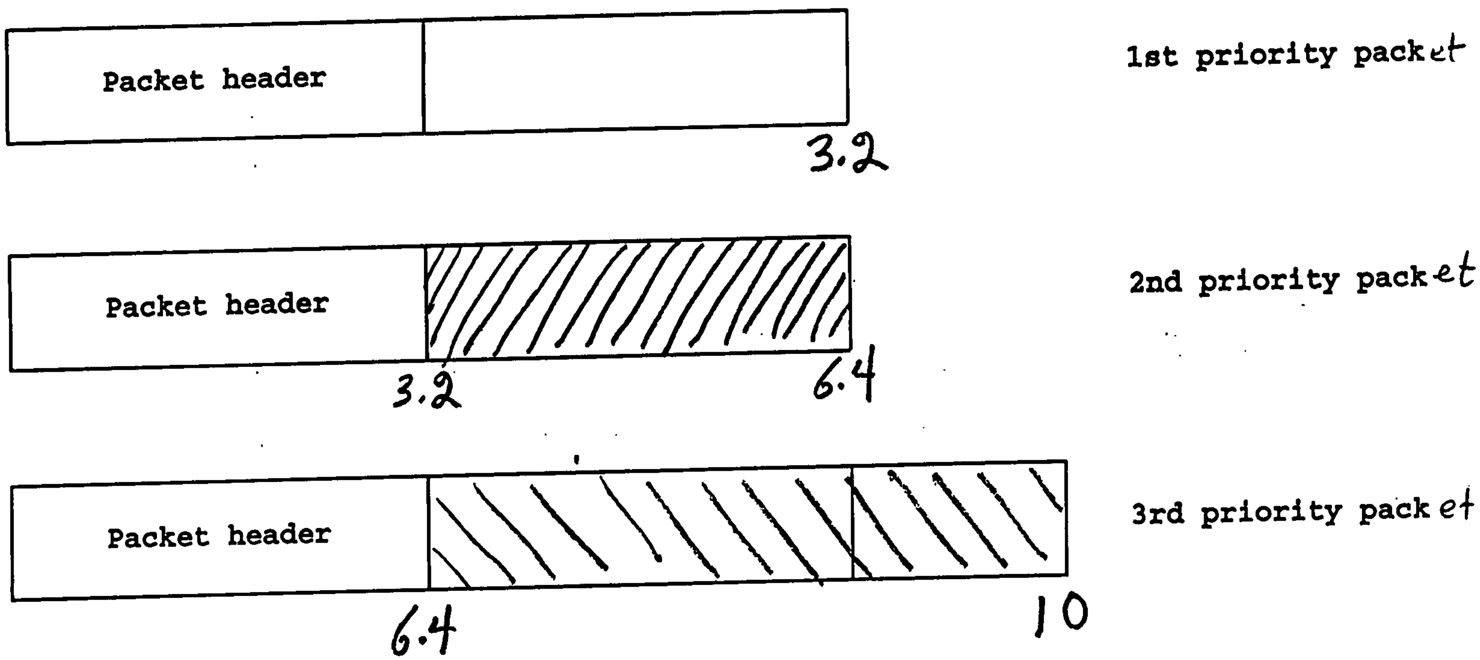


Fig. 4B

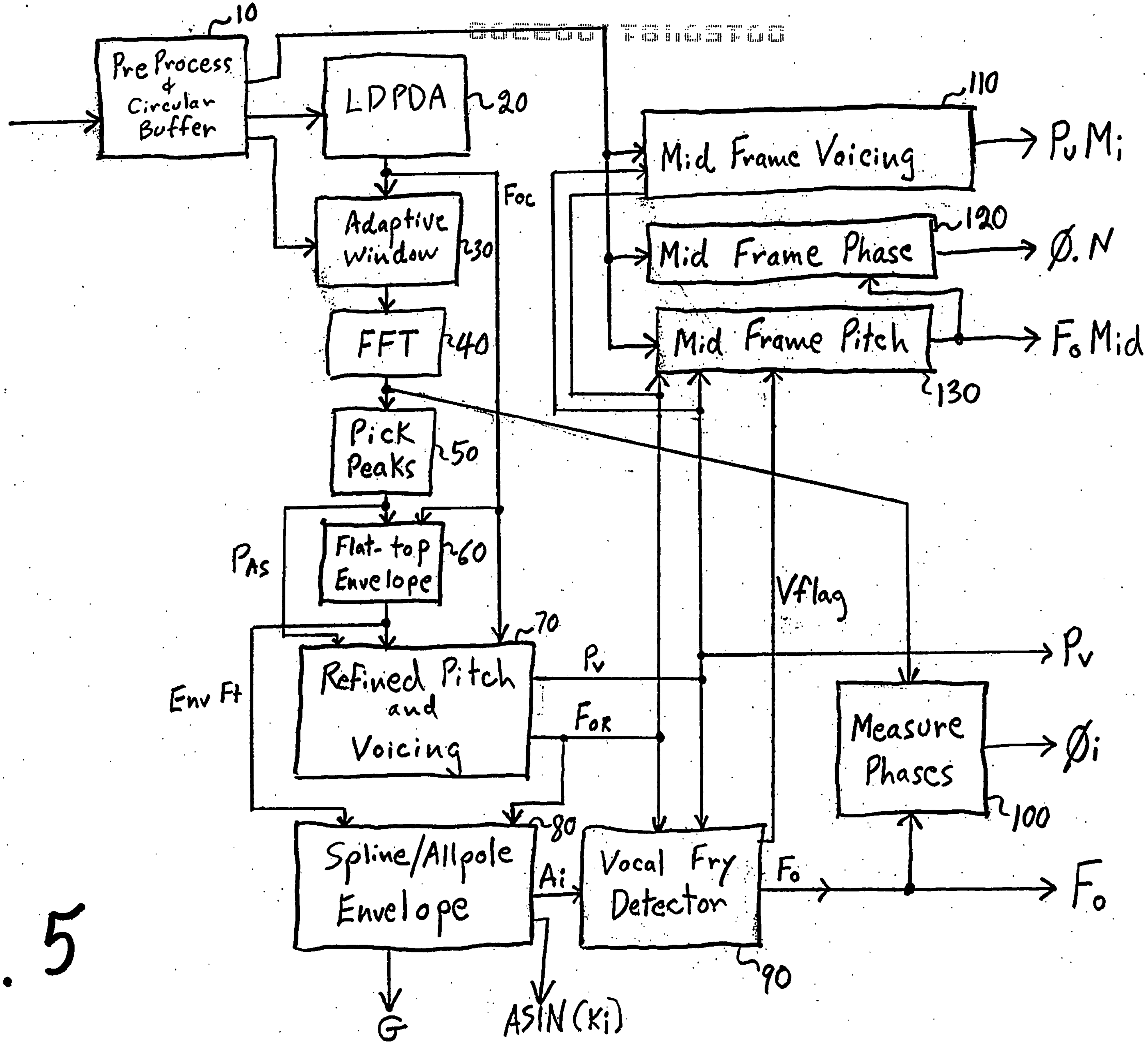
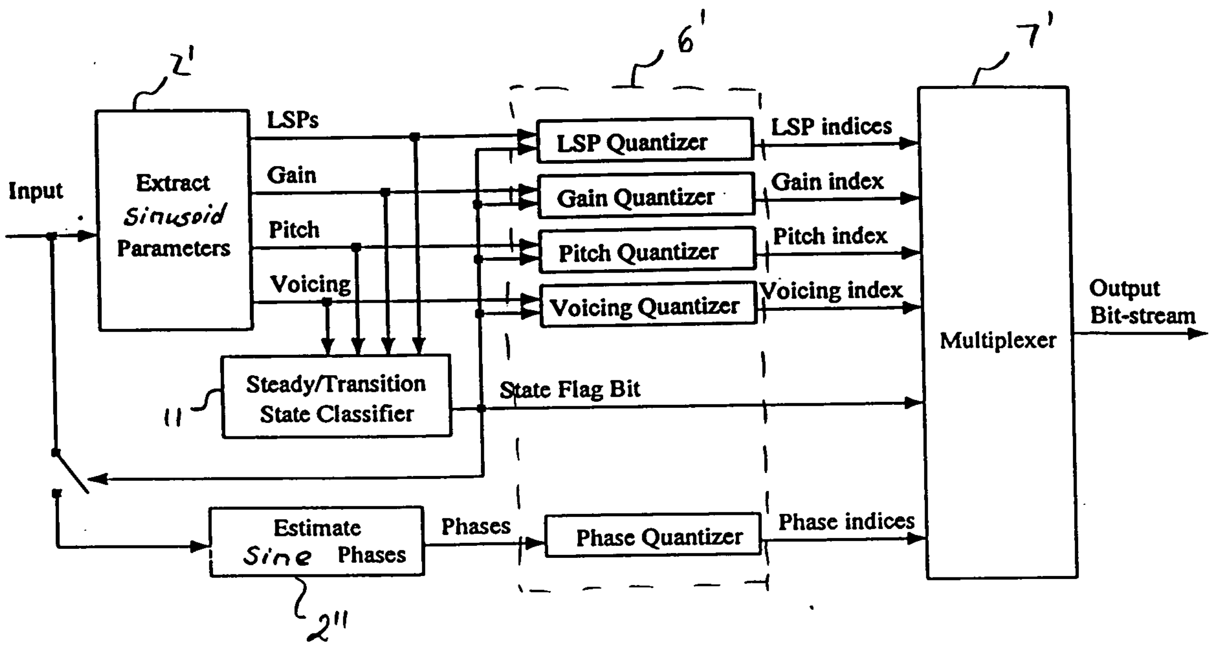

Pre Process & Circular Buffer
10
LDPDA
20
Adaptive Window
30
FFT
40
Pick Peaks
50
Flat-top Envelope
60
Refined Pitch and Voicing
70
Spline/Allpole Envelope
80
Vocal Fry Detector
90
Measure Phases
100
Mid Frame Voicing
110
Mid Frame Phase
120
Mid Frame Pitch
130
Foc
PAS
Env Ft
Pv
FoR
Ai
Fo
Vflag
Pv Mi
Ø. N
Fo Mid
Pv
Øi
Fo
G
ASIN (Ki)

Fig. 5



2'
6'
7'
Input
Extract
Sinusoid
Parameters
LSPs
Gain
Pitch
Voicing
LSP Quantizer
Gain Quantizer
Pitch Quantizer
Voicing Quantizer
LSP indices
Gain index
Pitch index
Voicing index
Multiplexer
Output
Bit-stream
Steady/Transition
State Classifier
11
State Flag Bit
Estimate
Sine Phases
Phases
Phase Quantizer
Phase indices
2''

Fig 5A

SPEECH SYNTHESIZER

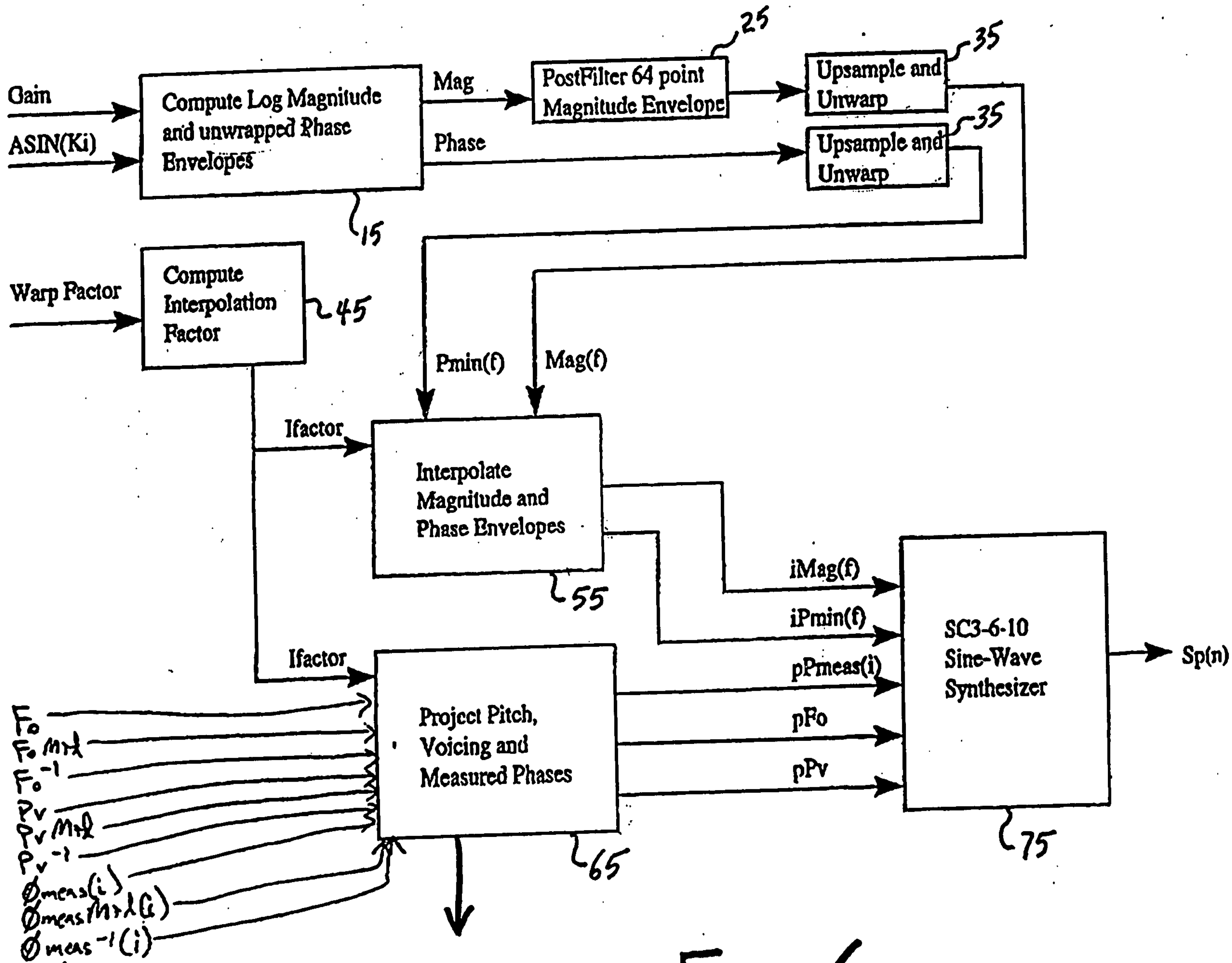
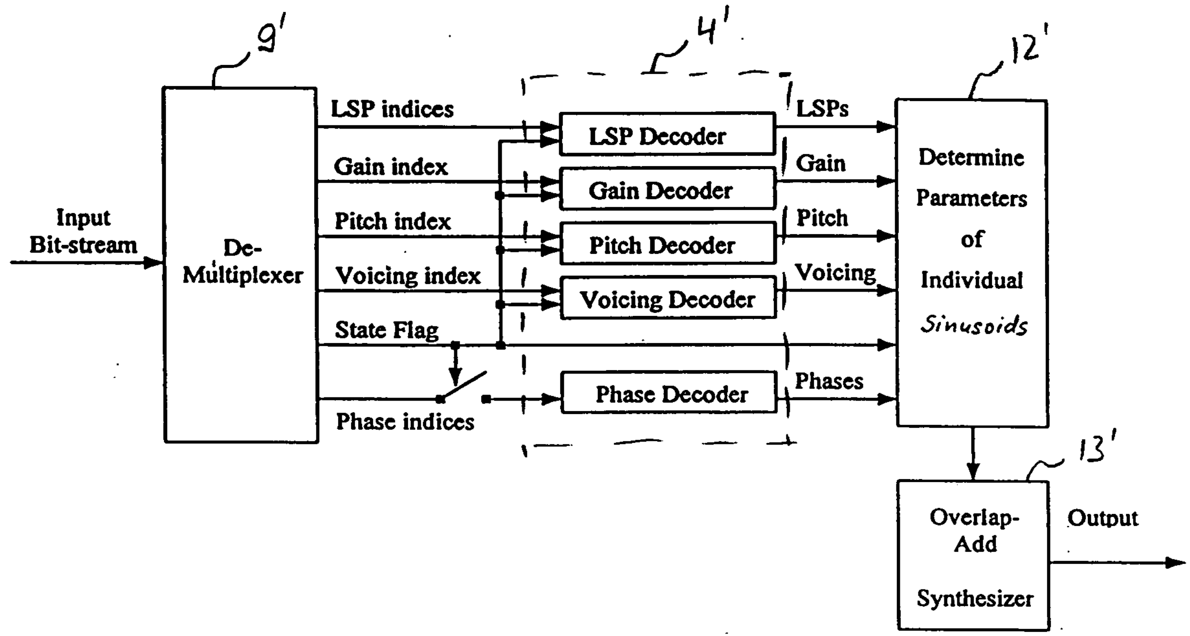


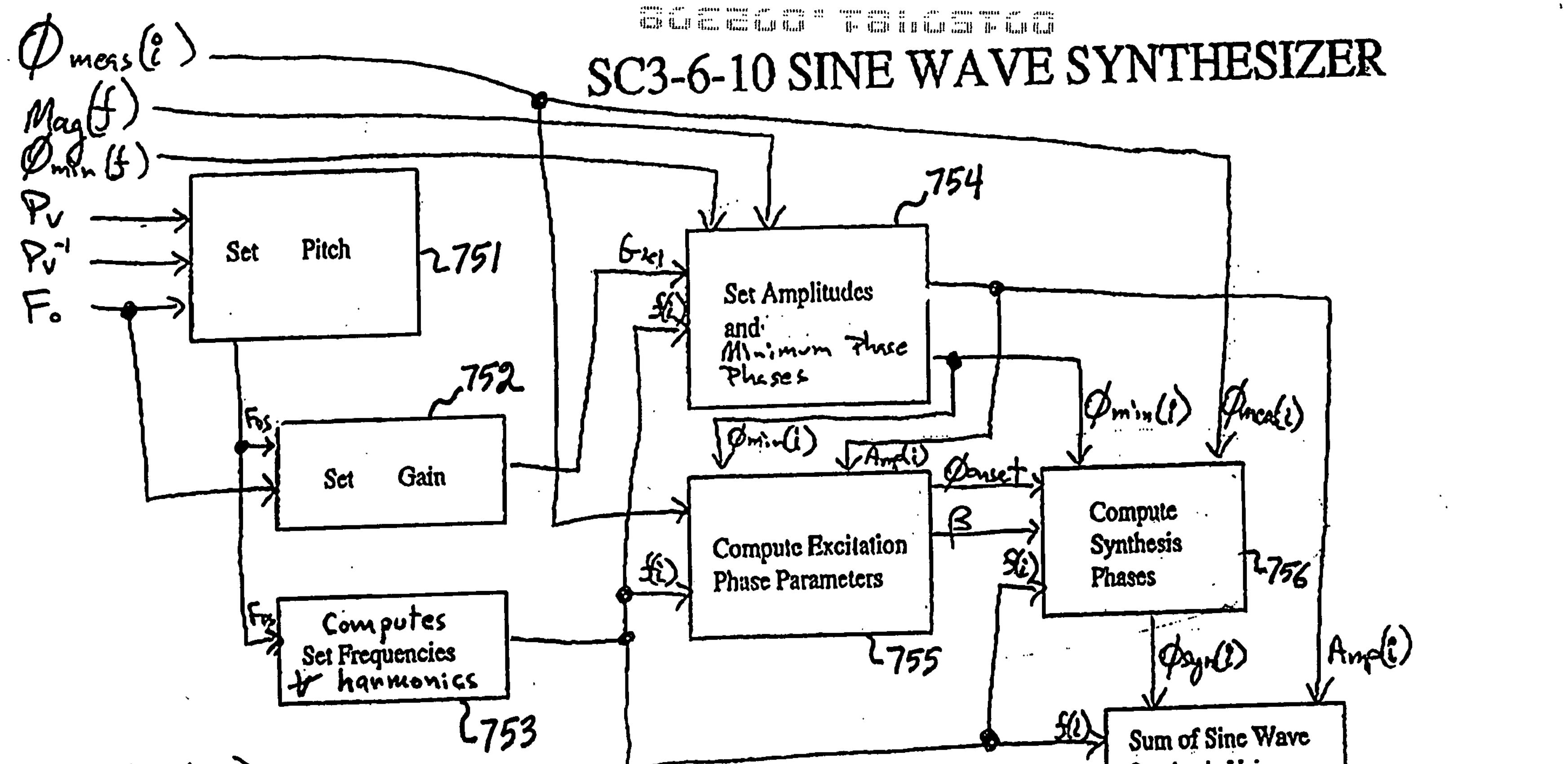
Fig. 6



9'
4'
12'
Input
Bit-stream
De-
Multiplexer
LSP indices
Gain index
Pitch index
Voicing index
State Flag
Phase indices
LSP Decoder
Gain Decoder
Pitch Decoder
Voicing Decoder
Phase Decoder
LSPs
Gain
Pitch
Voicing
Phases
Determine
Parameters
of
Individual
Sinusoids
13'
Overlap-
Add
Synthesizer
Output

Fig. 6A

SC3-6-10 SINE WAVE SYNTHESIZER



F_o pitch (input)
F_{os} pitch (synthesis)
P_v voicing probability
P_v^{-1} previous voicing probability
$Mag(f)$ magnitude spectrum
$\phi_{min}(f)$ minimum phase spectrum
$\phi_{meas}(i)$ measured phases
$f(i)$ synthesis frequencies
$Amp(i)$ synthesis amplitudes
$\phi_{syn}(i)$ synthesis phases
$\phi_{min}(i)$ minimum phases at $f(i)$

Fig. 7

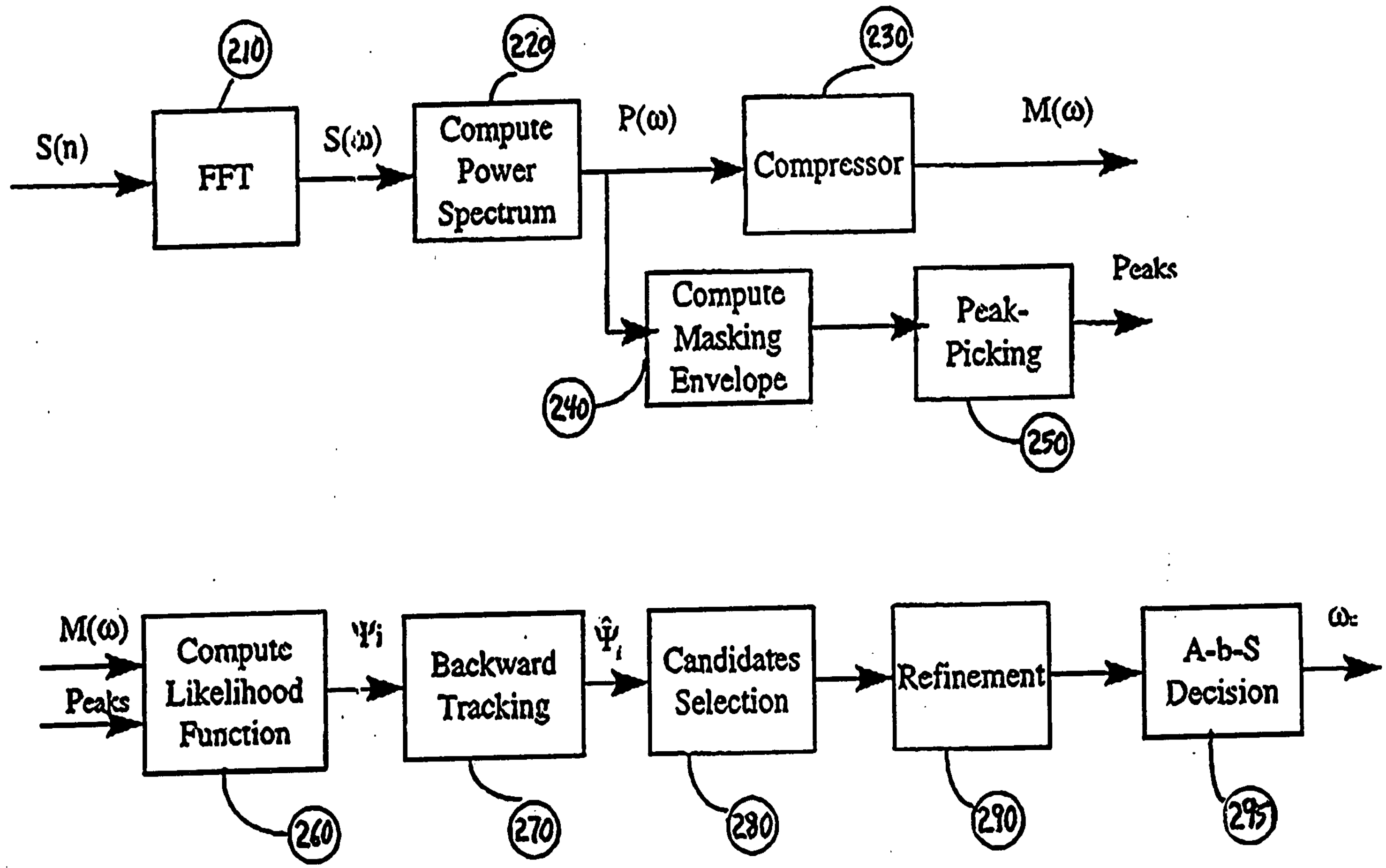

S(n)
FFT
210
S(ω)
Compute Power Spectrum
220
P(ω)
Compressor
230
M(ω)
Compute Masking Envelope
240
Peak-Picking
250
Peaks
M(ω)
Peaks
Compute Likelihood Function
260
Ψi
Backward Tracking
270
Ψ̂i
Candidates Selection
280
Refinement
290
A-b-S Decision
295
ωc

Fig. 8

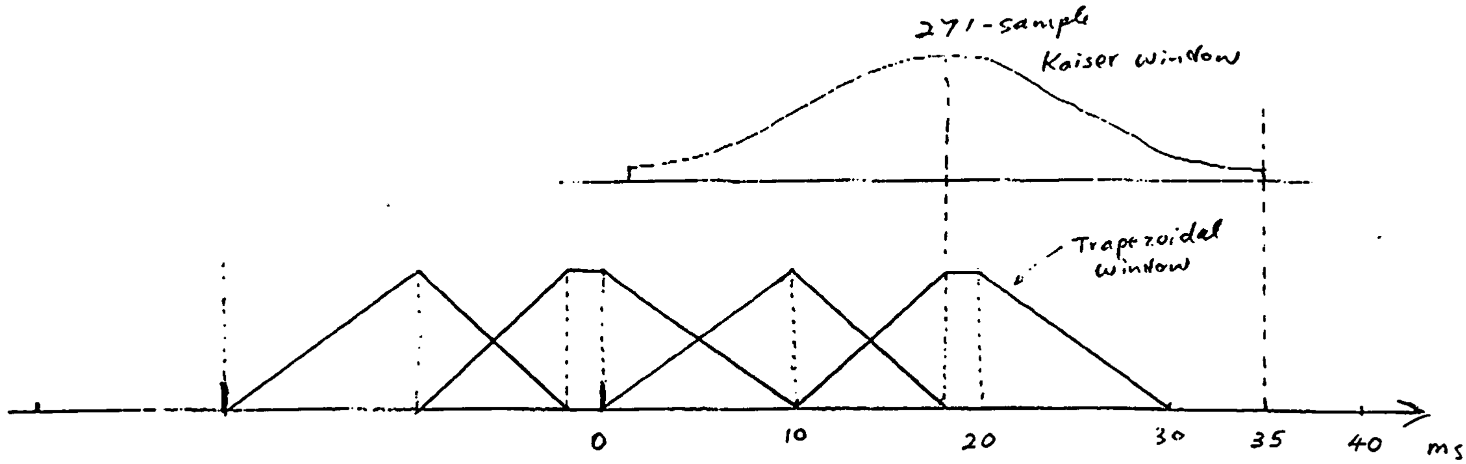

271-sample
Kaiser window
Trapezoidal
window
0
10
20
30
35
40
ms

Fig. 8. A

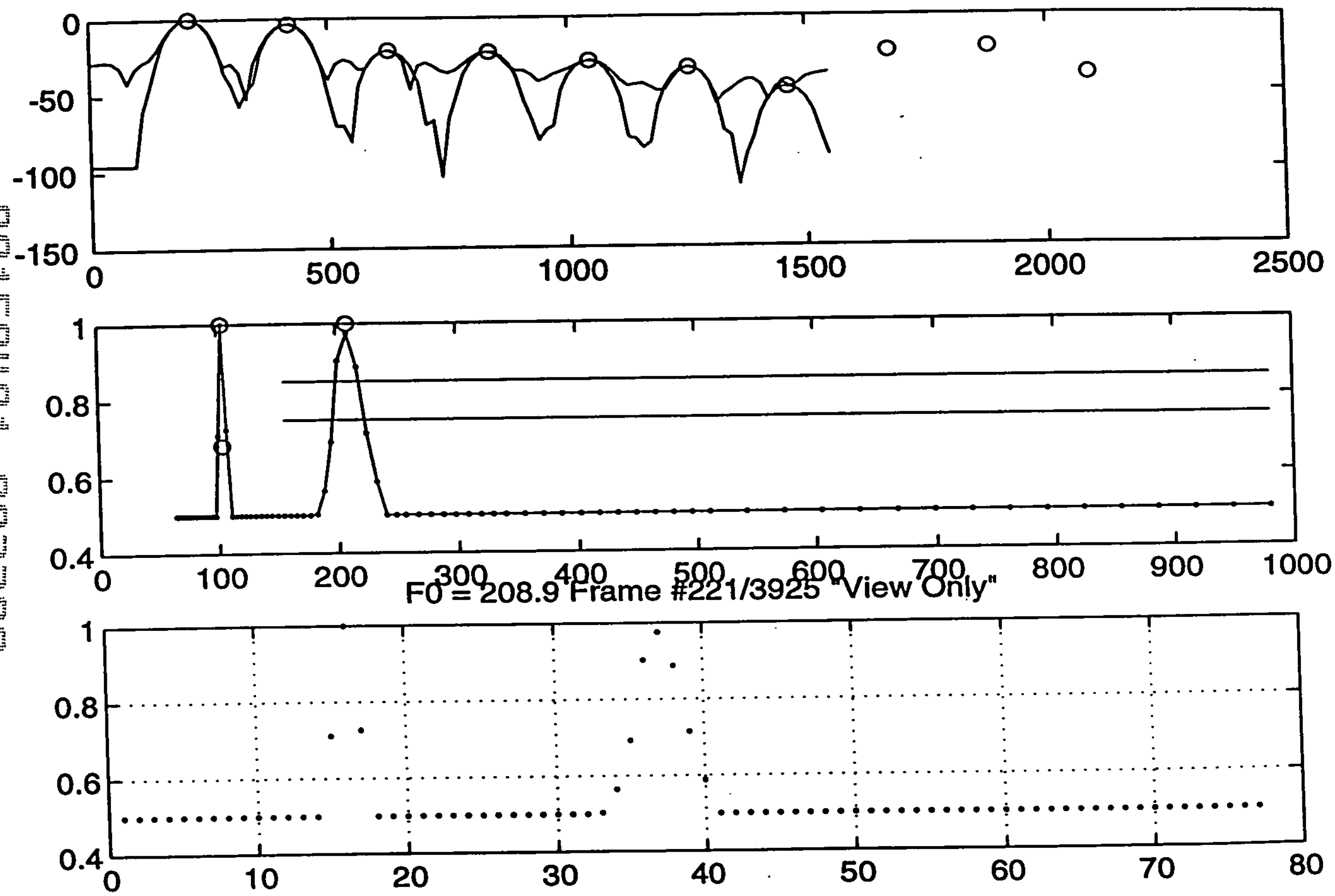

0
-50
-100
-150
0
500
1000
1500
2000
2500
1
0.8
0.6
0.4
100
200
300
400
500
600
700
800
900
1000
F0 = 208.9 Frame #221/3925 "View Only"
1
0.8
0.6
0.4
0
10
20
30
40
50
60
70
80

Fig. 9A

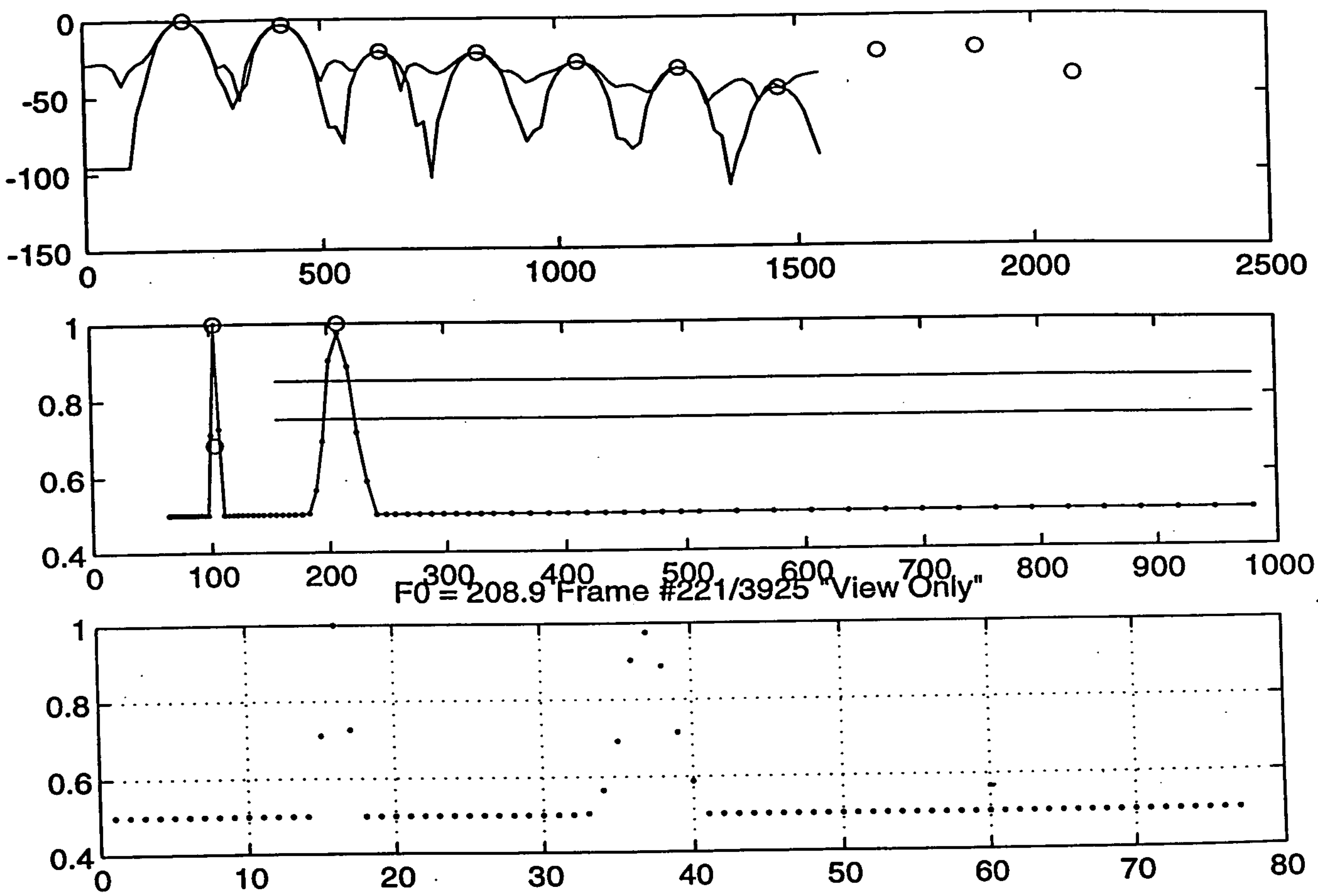
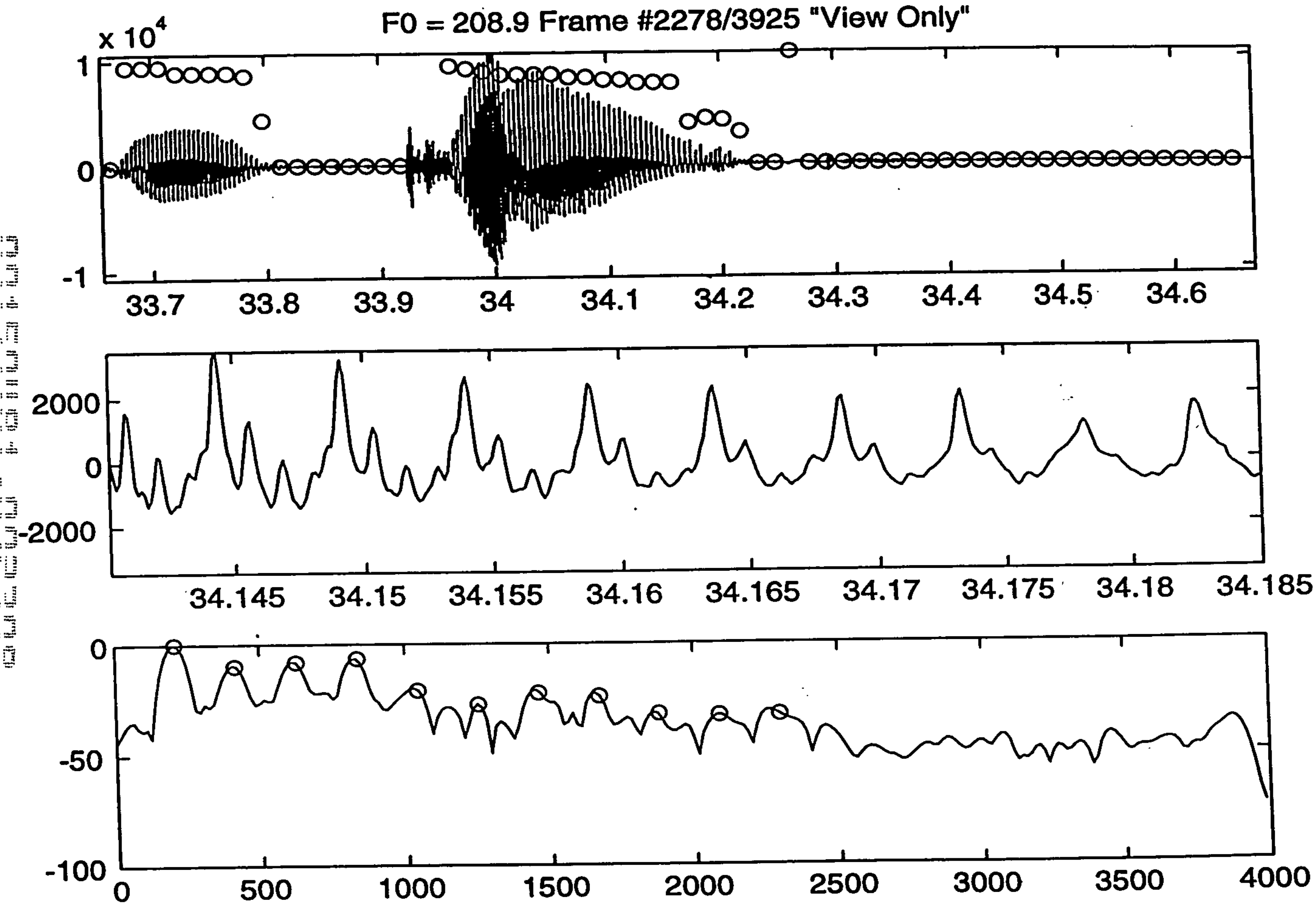

0
-50
-100
-150
0
500
1000
1500
2000
2500
1
0.8
0.6
0.4
0
100
200
300
400
500
600
700
800
900
1000
F0 = 208.9 Frame #221/3925 "View Only"
1
0.8
0.6
0.4
0
10
20
30
40
50
60
70
80

Fig. 9B



F0 = 208.9 Frame #2278/3925 "View Only"
x 10^4
1
0
-1
33.7
33.8
33.9
34
34.1
34.2
34.3
34.4
34.5
34.6
2000
0
-2000
34.145
34.15
34.155
34.16
34.165
34.17
34.175
34.18
34.185
0
-50
-100
0
500
1000
1500
2000
2500
3000
3500
4000

Fig. 9C

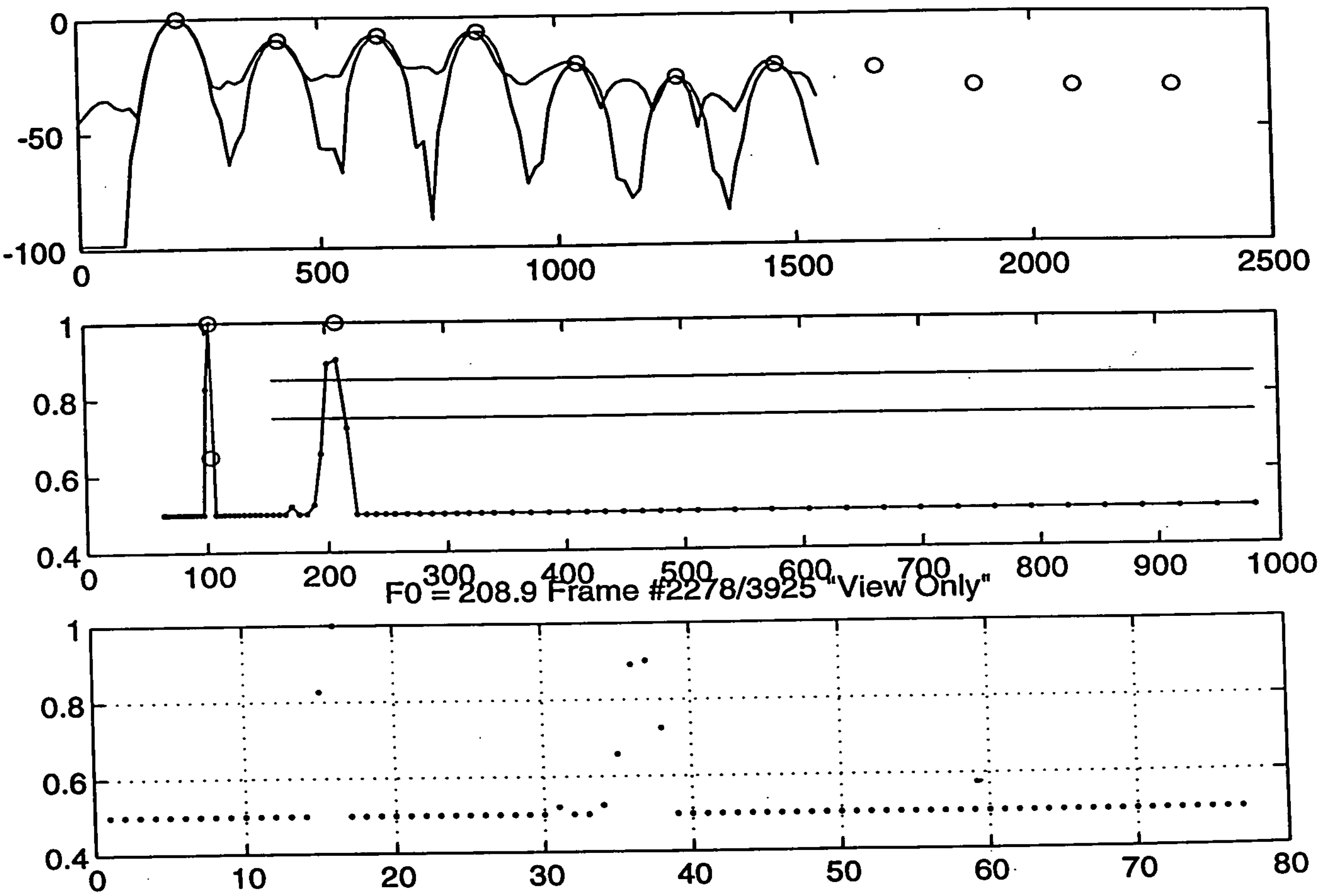
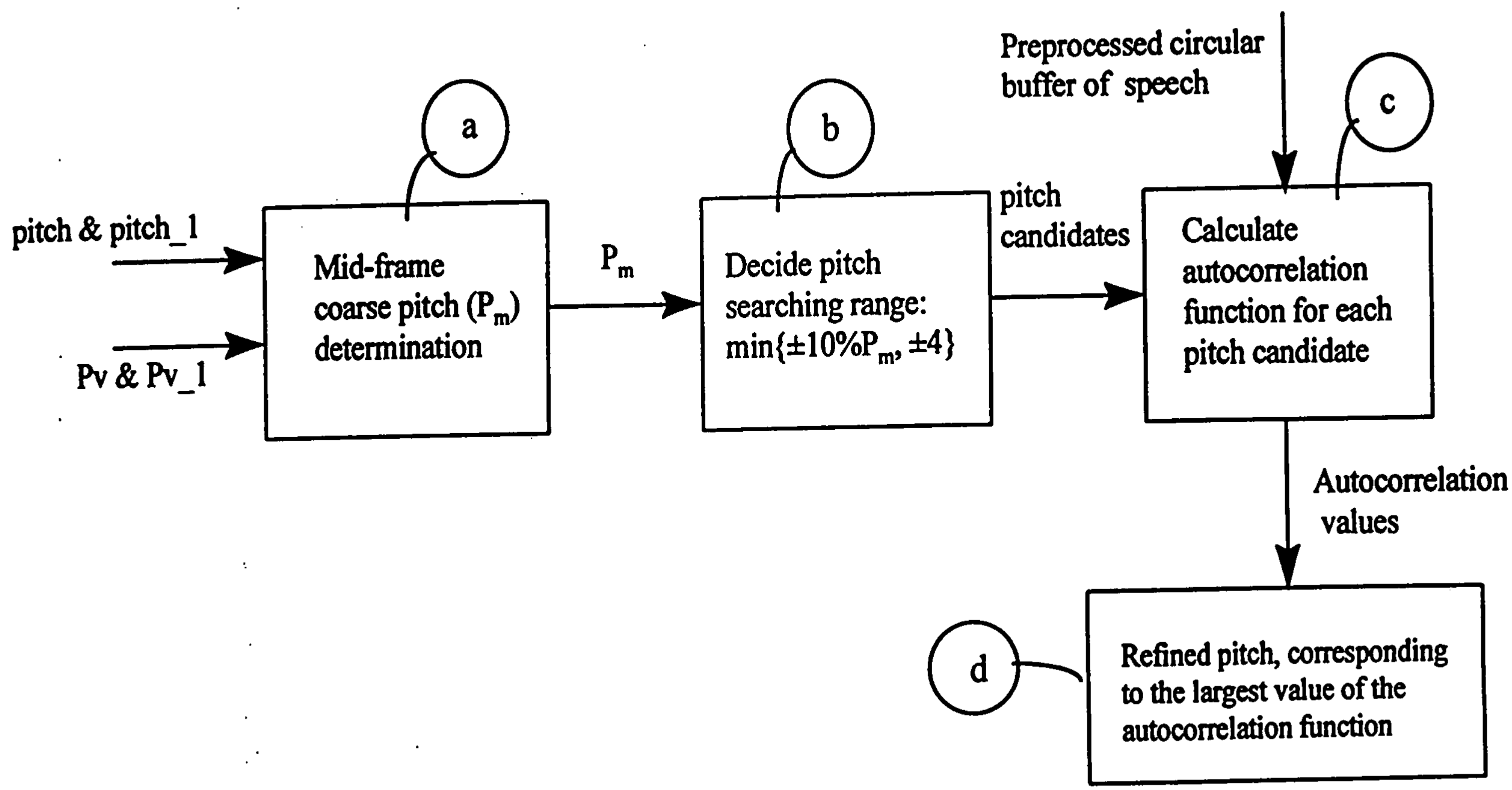

0
-50
-100
0
500
1000
1500
2000
2500
1
0.8
0.6
0.4
0
100
200
300
400
500
600
700
800
900
1000
F0 = 208.9 Frame #2278/3925 "View Only"
1
0.8
0.6
0.4
0
10
20
30
40
50
60
70
80

Fig. 9D



pitch & pitch_1
Pv & Pv_1
a
Mid-frame coarse pitch (P_m) determination
P_m
b
Decide pitch searching range: min{±10%P_m, ±4}
pitch candidates
Preprocessed circular buffer of speech
c
Calculate autocorrelation function for each pitch candidate
Autocorrelation values
d
Refined pitch, corresponding to the largest value of the autocorrelation function

Fig. 10

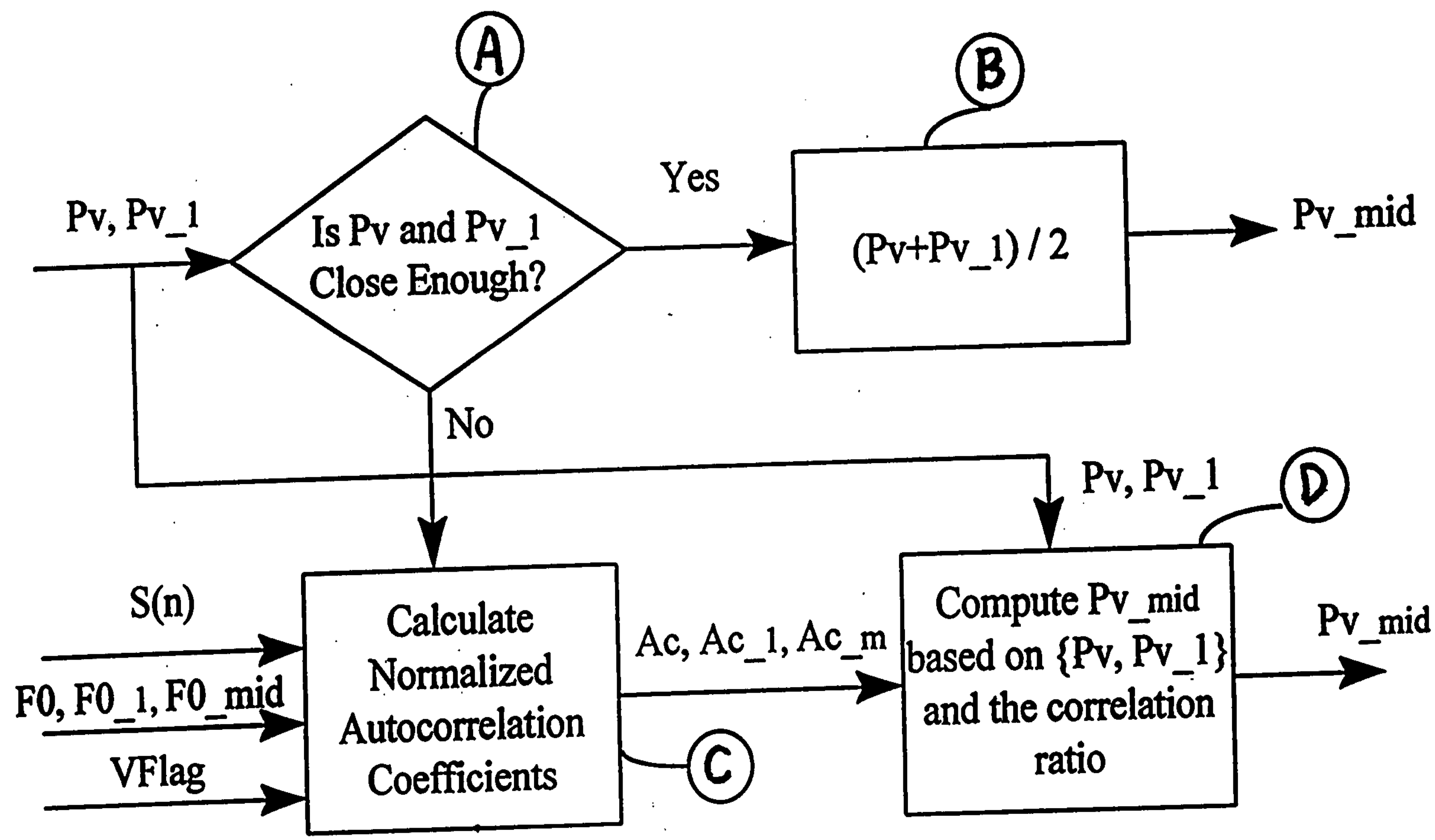
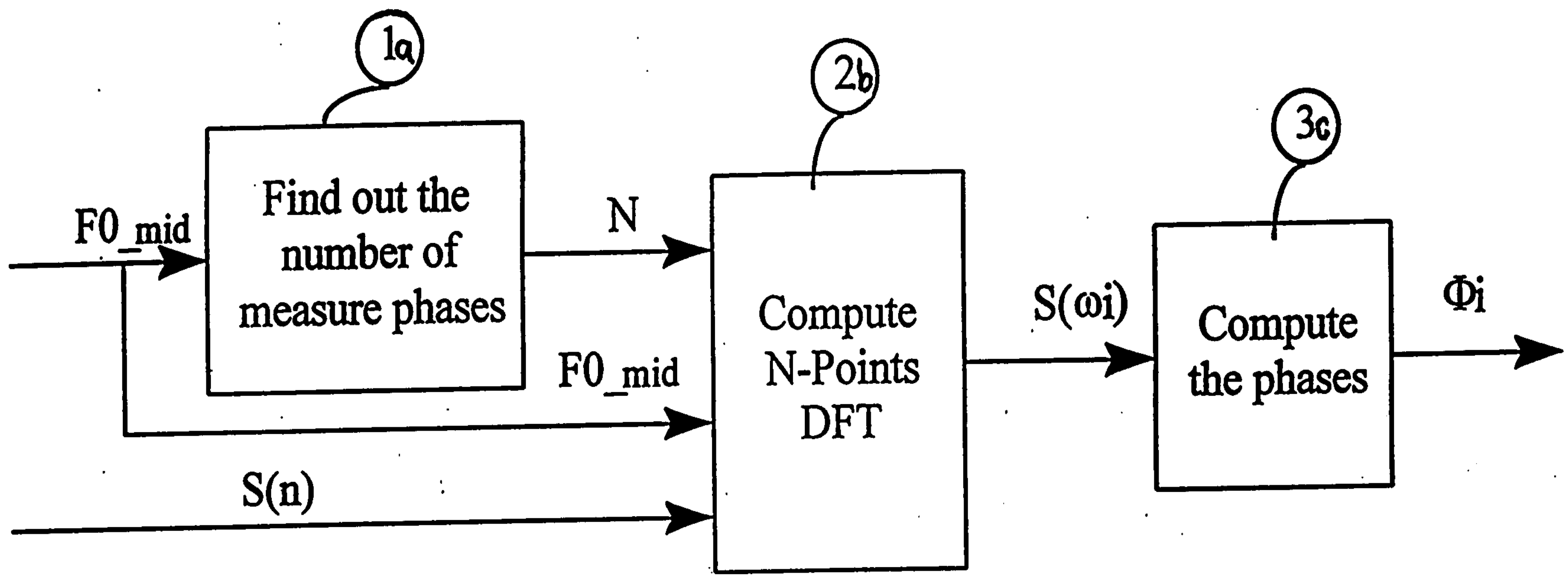

Pv, Pv_1
Is Pv and Pv_1 Close Enough?
A
Yes
No
B
(Pv+Pv_1) / 2
Pv_mid
S(n)
F0, F0_1, F0_mid
VFlag
Calculate Normalized Autocorrelation Coefficients
C
Ac, Ac_1, Ac_m
Pv, Pv_1
D
Compute Pv_mid based on {Pv, Pv_1} and the correlation ratio
Pv_mid

Fig. 11



1a
Find out the number of measure phases
F0_mid
N
F0_mid
S(n)
2b
Compute N-Points DFT
S(ωi)
3c
Compute the phases
Φi

Fig. 12

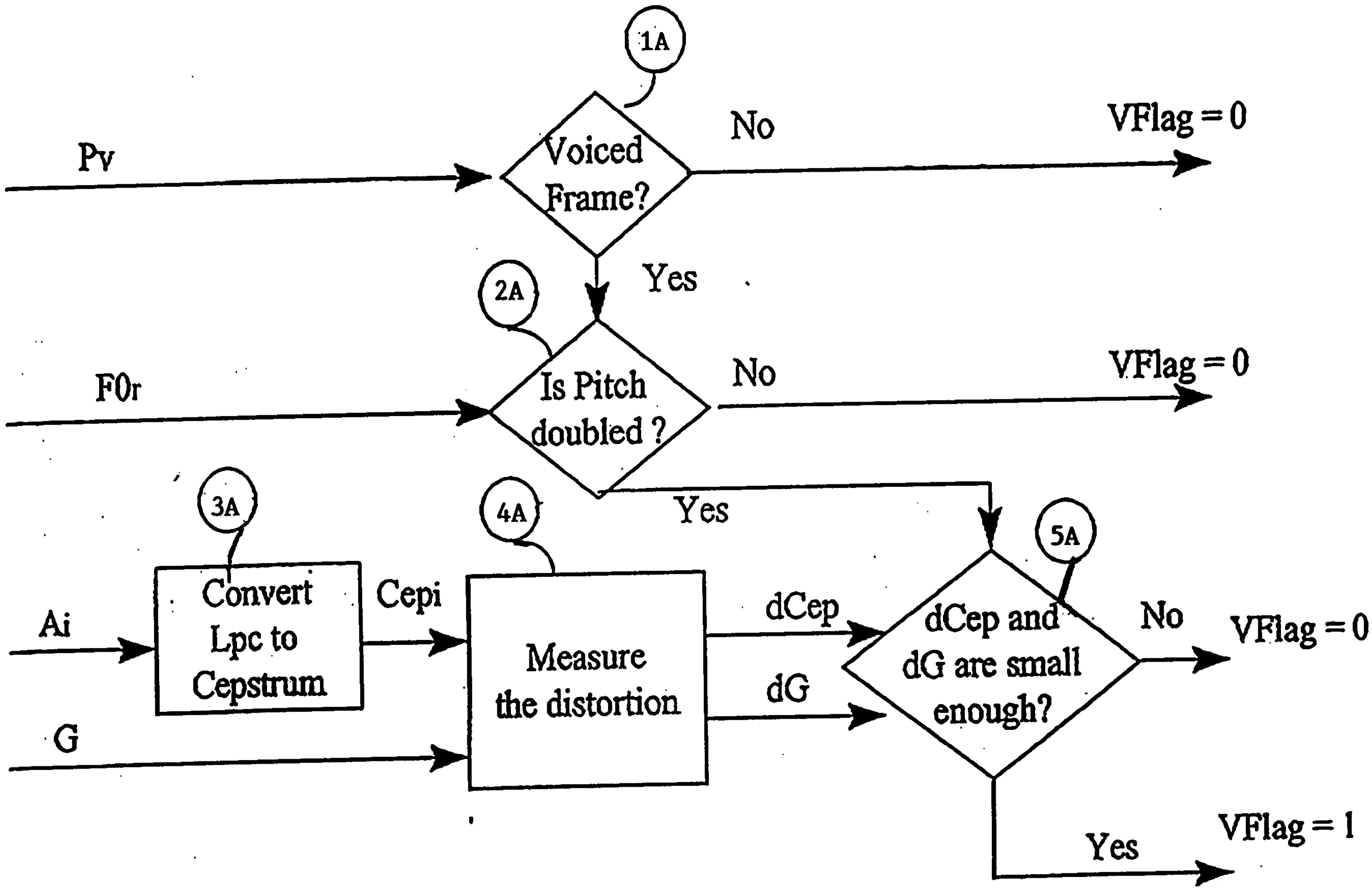

1A
Pv
Voiced Frame?
No
VFlag = 0
Yes
2A
F0r
Is Pitch doubled ?
No
VFlag = 0
Yes
3A
Ai
Convert Lpc to Cepstrum
Cepi
4A
G
Measure the distortion
dCep
dG
5A
dCep and dG are small enough?
No
VFlag = 0
Yes
VFlag = 1

Fig. 13

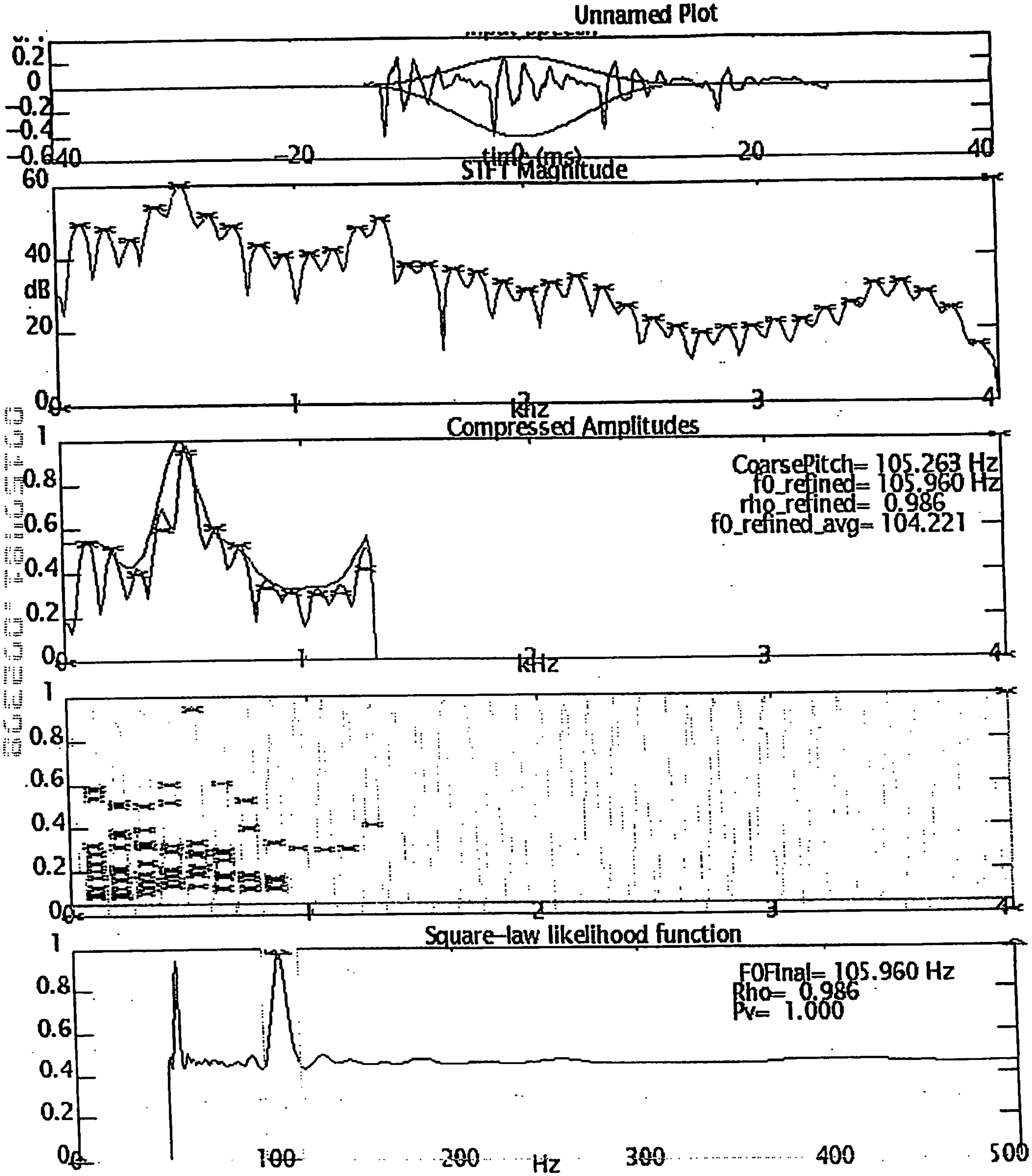
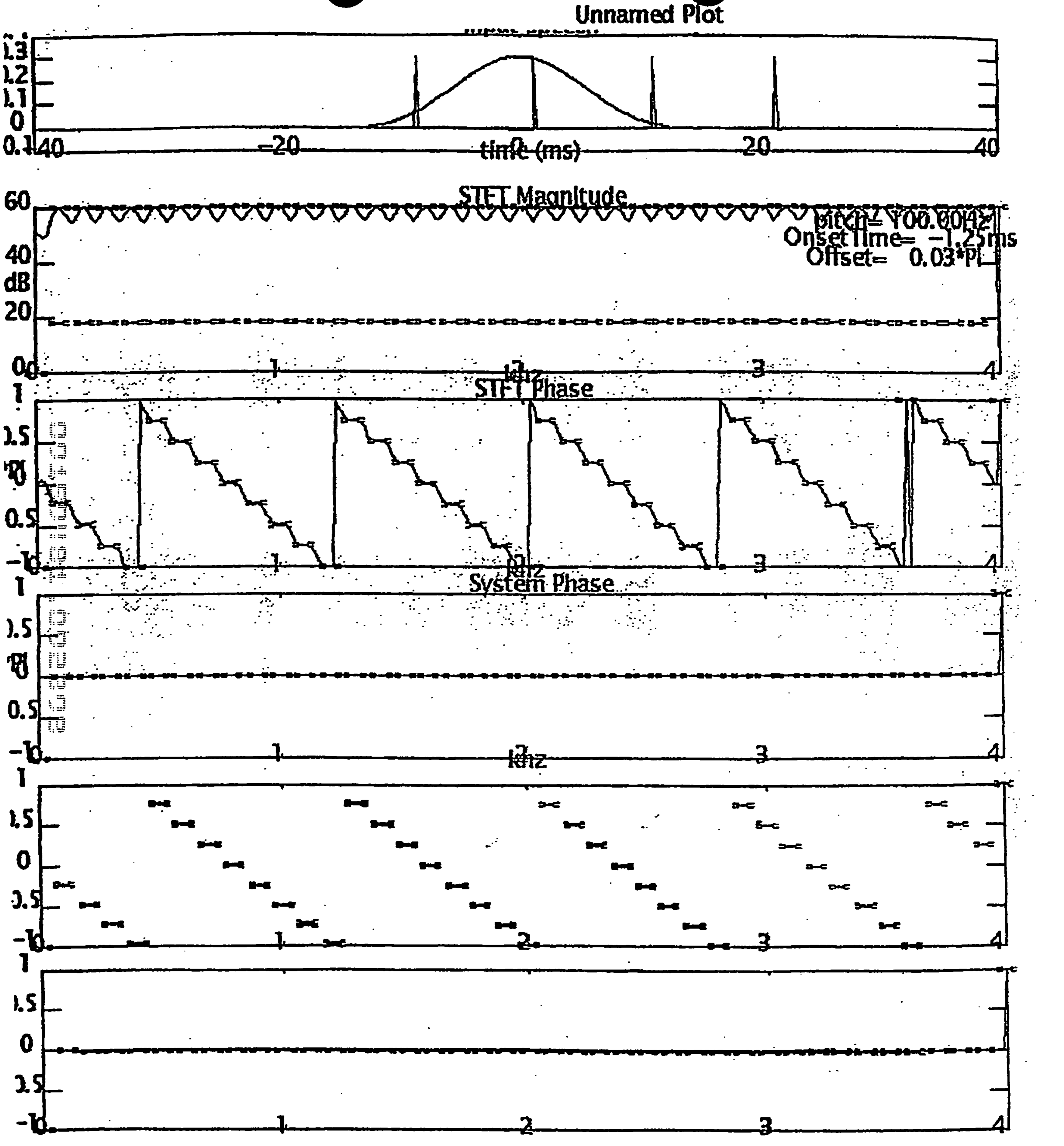
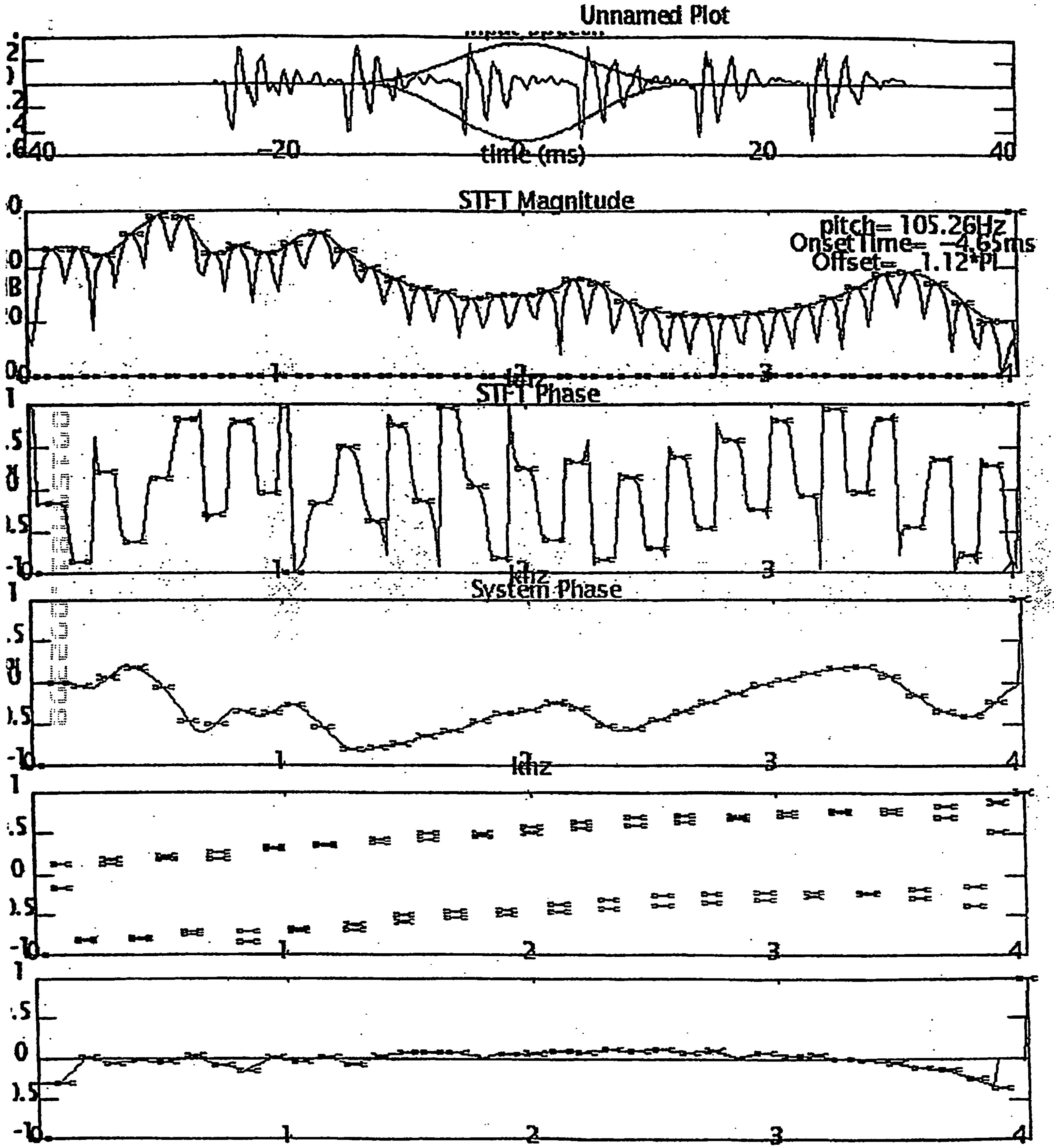

Unnamed Plot
0.2
0
-0.2
-0.4
-0.6
-40
-20
time (ms)
20
40
STFT Magnitude
60
40
dB
20
0
1
2
kHz
3
4
Compressed Amplitudes
CoarsePitch= 105.263 Hz
f0_refined= 105.960 Hz
rho_refined= 0.986
f0_refined_avg= 104.221
1
0.8
0.6
0.4
0.2
0
kHz
Square-law likelihood function
F0Final= 105.960 Hz
Rho= 0.986
Pv= 1.000
100
200
Hz
300
400
500

Fig. 14



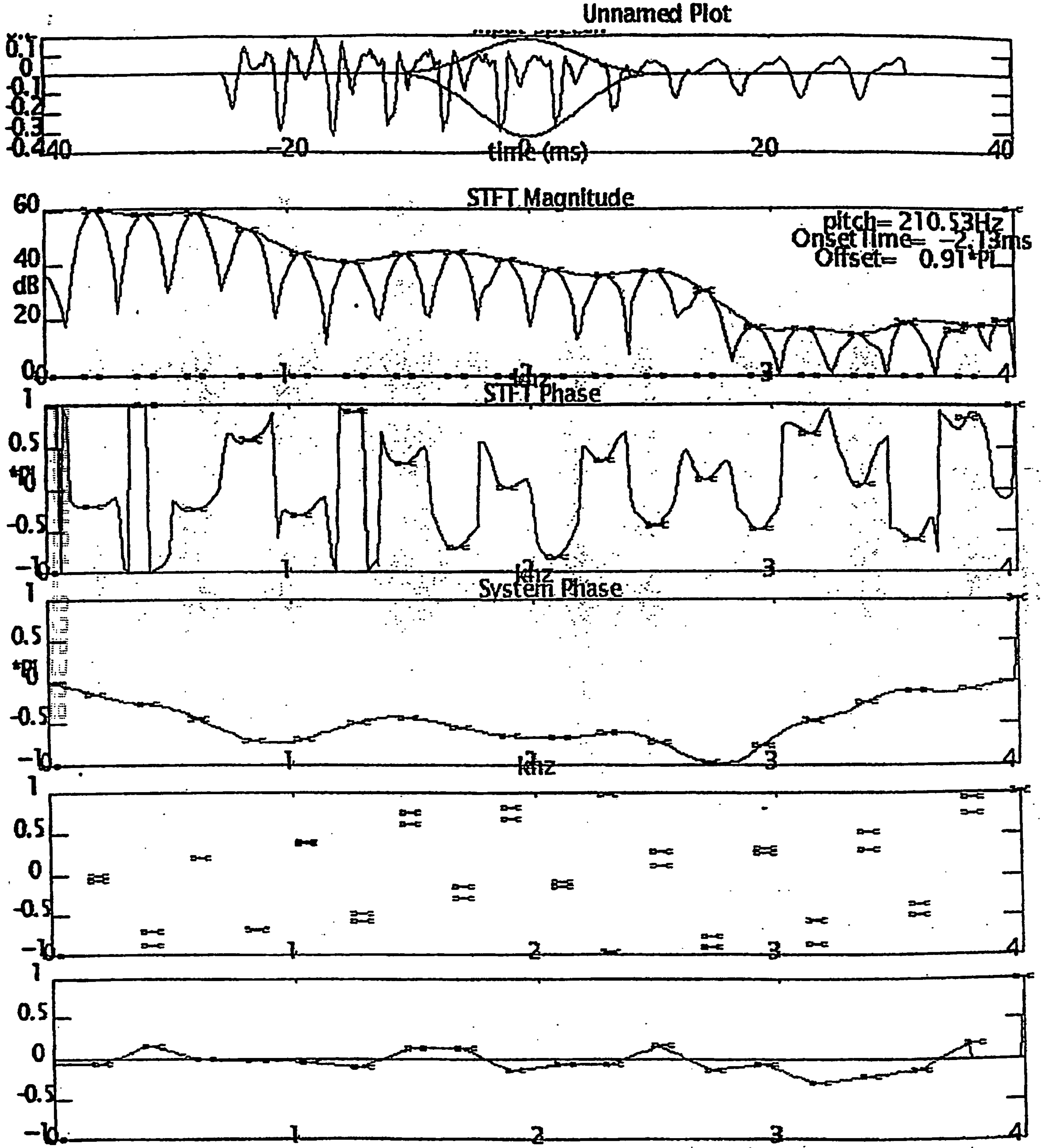
Unnamed Plot
time (ms)
STFT Magnitude
Pitch= 100.00Hz
Onset Time= -1.25ms
Offset= 0.03*PI
dB
kHz
STFT Phase
PI
System Phase

Fig. 15



Unnamed Plot
time (ms)
STFT Magnitude
pitch= 105.26Hz
OnsetTime= -4.65ms
Offset= 1.12*PL
STFT Phase
khz
System Phase
khz

Fig. 16



Unnamed Plot
time (ms)
STFT Magnitude
pitch= 210.53Hz
Onset Time= -2.13ms
Offset= 0.91*PI
dB
khz
STFT Phase
*PI
System Phase

Fig. 17

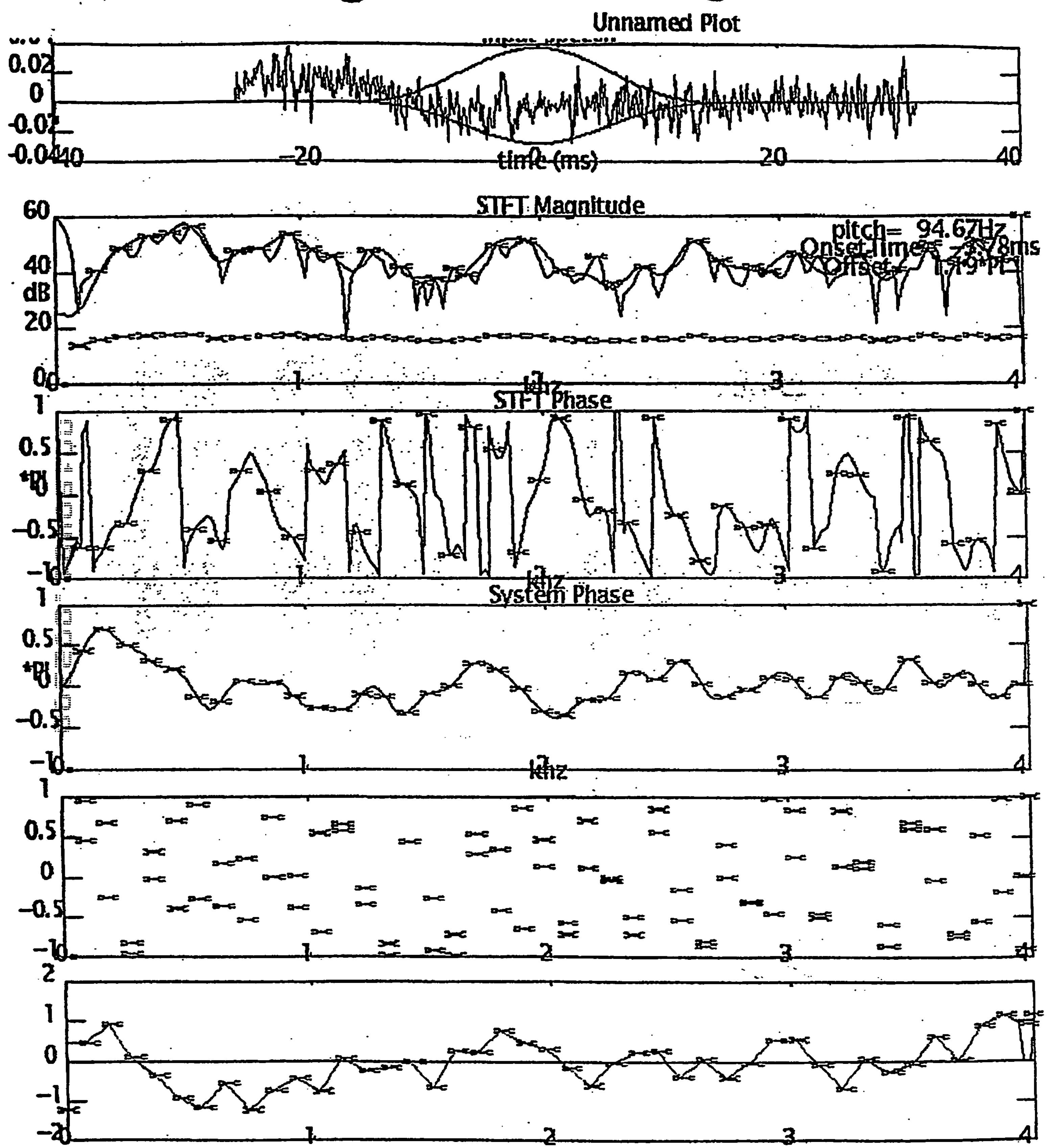
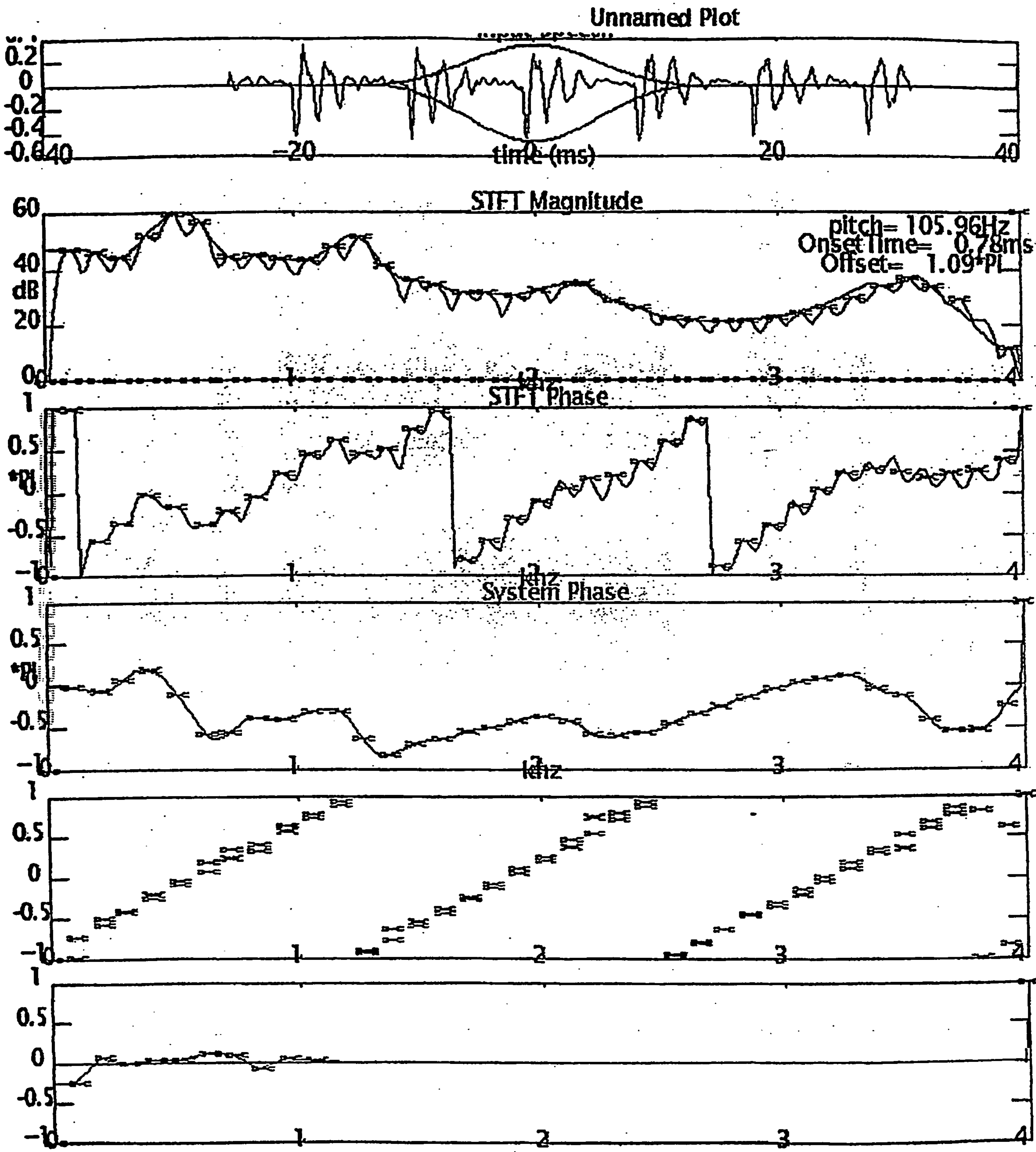

Unnamed Plot
time (ms)
STFT Magnitude
pitch= 94.67Hz
dB
khz
STFT Phase
*PI
System Phase

Fig. 18



Unnamed Plot
time (ms)
STFT Magnitude
pitch= 105.96Hz
OnsetTime= 0.78ms
Offset= 1.09*PI
dB
kHz
STFT Phase
*PI
System Phase

Fig. 19

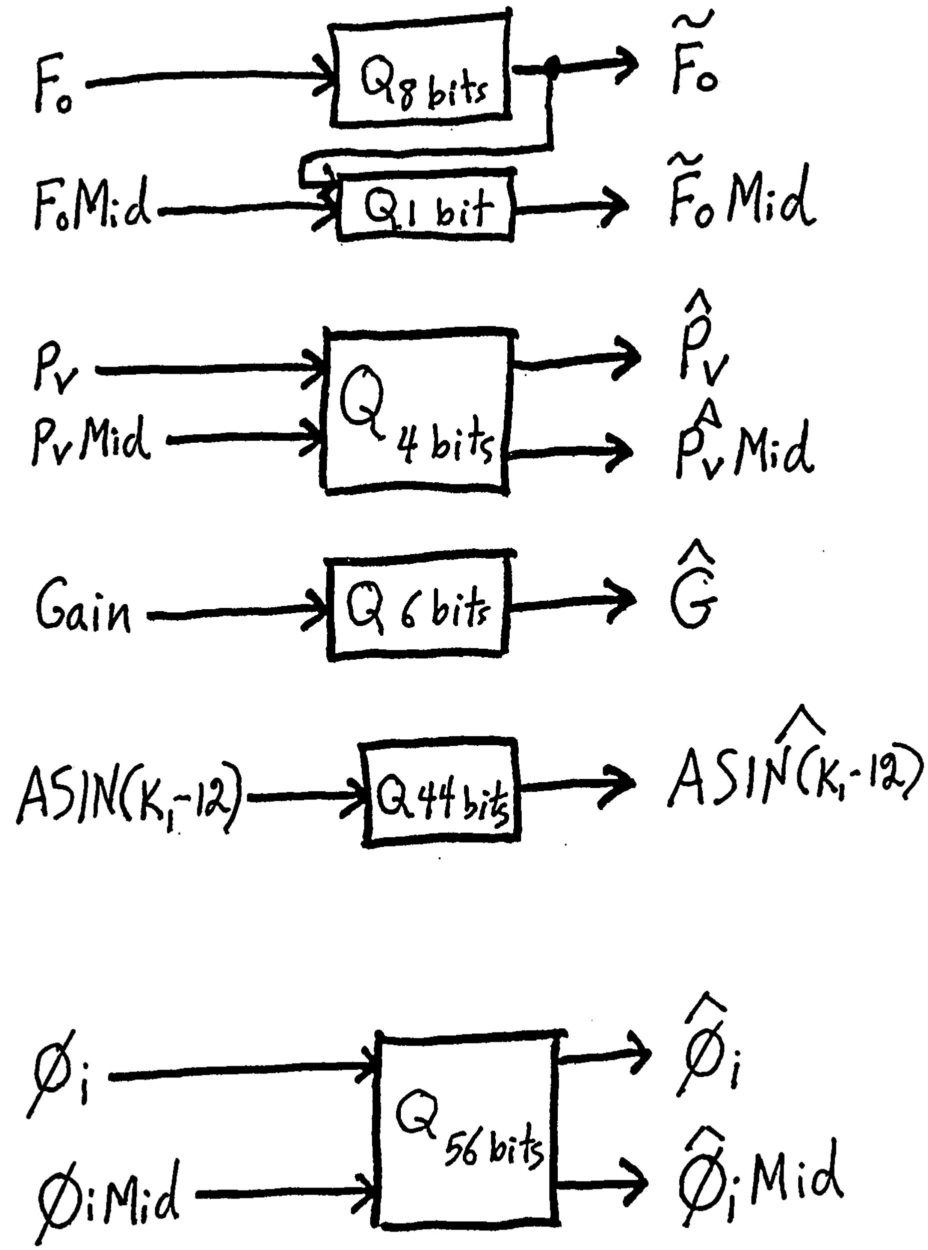

F0
F0Mid
Q8 bits
Q1 bit
F̃0
F̃0 Mid
Pv
Pv Mid
Q 4 bits
P̂v
P̂v Mid
Gain
Q 6 bits
Ĝ
ASIN(K1-12)
Q 44 bits
ASIN(K1-12)
Øi
ØiMid
Q 56 bits
Ø̂i
Ø̂i Mid

Fig. 20

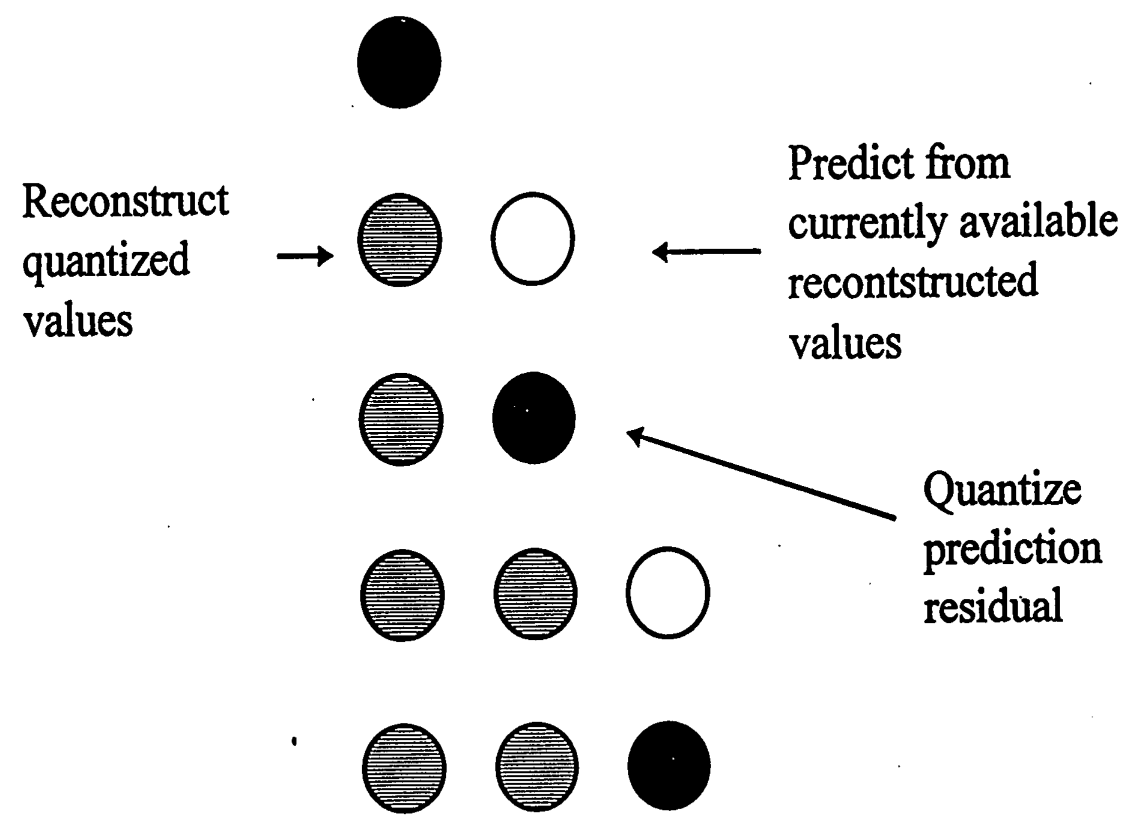

Reconstruct quantized values
Predict from currently available recontstructed values
Quantize prediction residual

Fig. 21

Spectral Quantization

prediction of AS(i)

$$\hat{AS}(i) = \sum_{j=1}^{i-1} a_{ij}\,\tilde{AS}(j)$$

$$\underbrace{\begin{bmatrix} \hat{AS}(1) \\ \hat{AS}(2) \\ \vdots \\ \hat{AS}(12) \end{bmatrix}}_{12\times 1} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ a_{21} & 0 & & & \\ a_{31} & a_{32} & \cancel{0} & & \\ \vdots & & & & \\ a_{12,1} & a_{12,2} & \cdots & & a_{12,11} \end{bmatrix}}_{12\times 11} \underbrace{\begin{bmatrix} \tilde{AS}(1) \\ \tilde{AS}(2) \\ \vdots \\ \tilde{AS}(11) \end{bmatrix}}_{11\times 1}$$

fixed stored

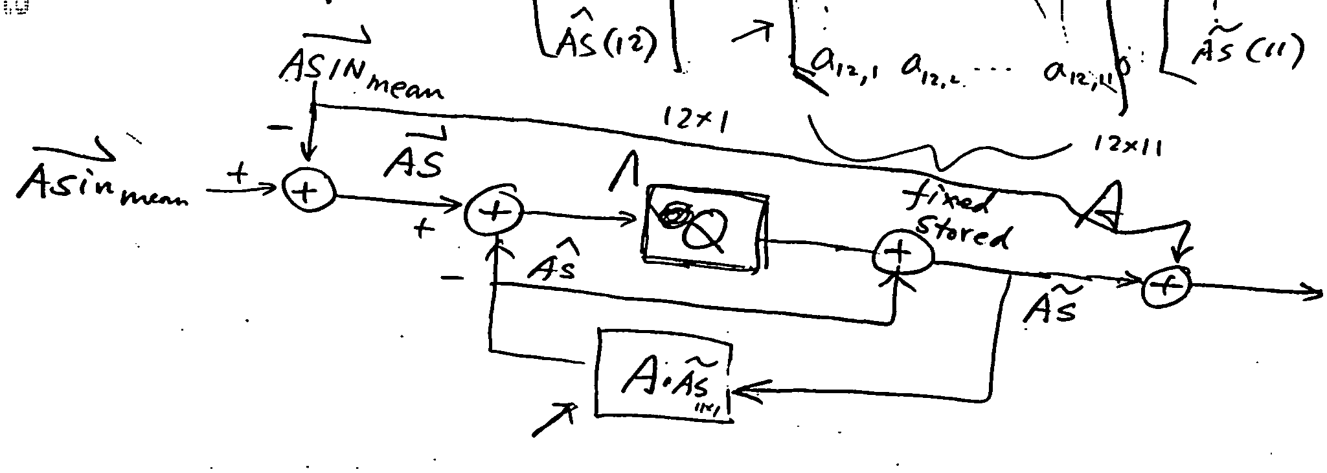
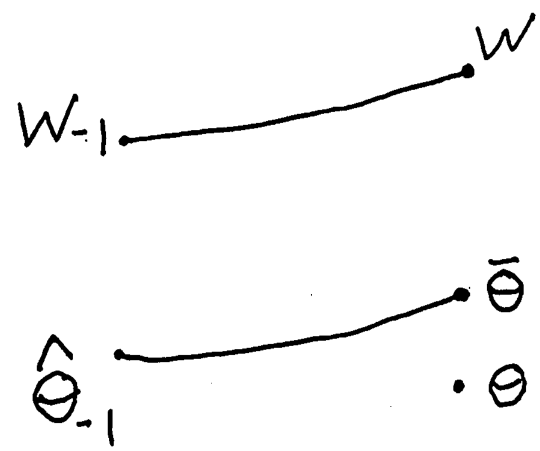


Fig. 21A

PHASE-PREDICTIVE CODING



$$\bar{\Theta} = \Theta_{-1} + \frac{W_{-1} + W}{2} \times T$$

PHASE RESIDUAL $= \Theta - \bar{\Theta}$

W_{-1} = frequency at previous frame
W = frequency at current frame
Θ_{-1} = quantized phase at previous frame
$\bar{\Theta}$ = predicted phase at current frame
Θ = measured phase at current frame

Fig. 22A

SCATTER PLOT OF 20ms PHASE AND 10ms PHASE PREDICTION ERROR

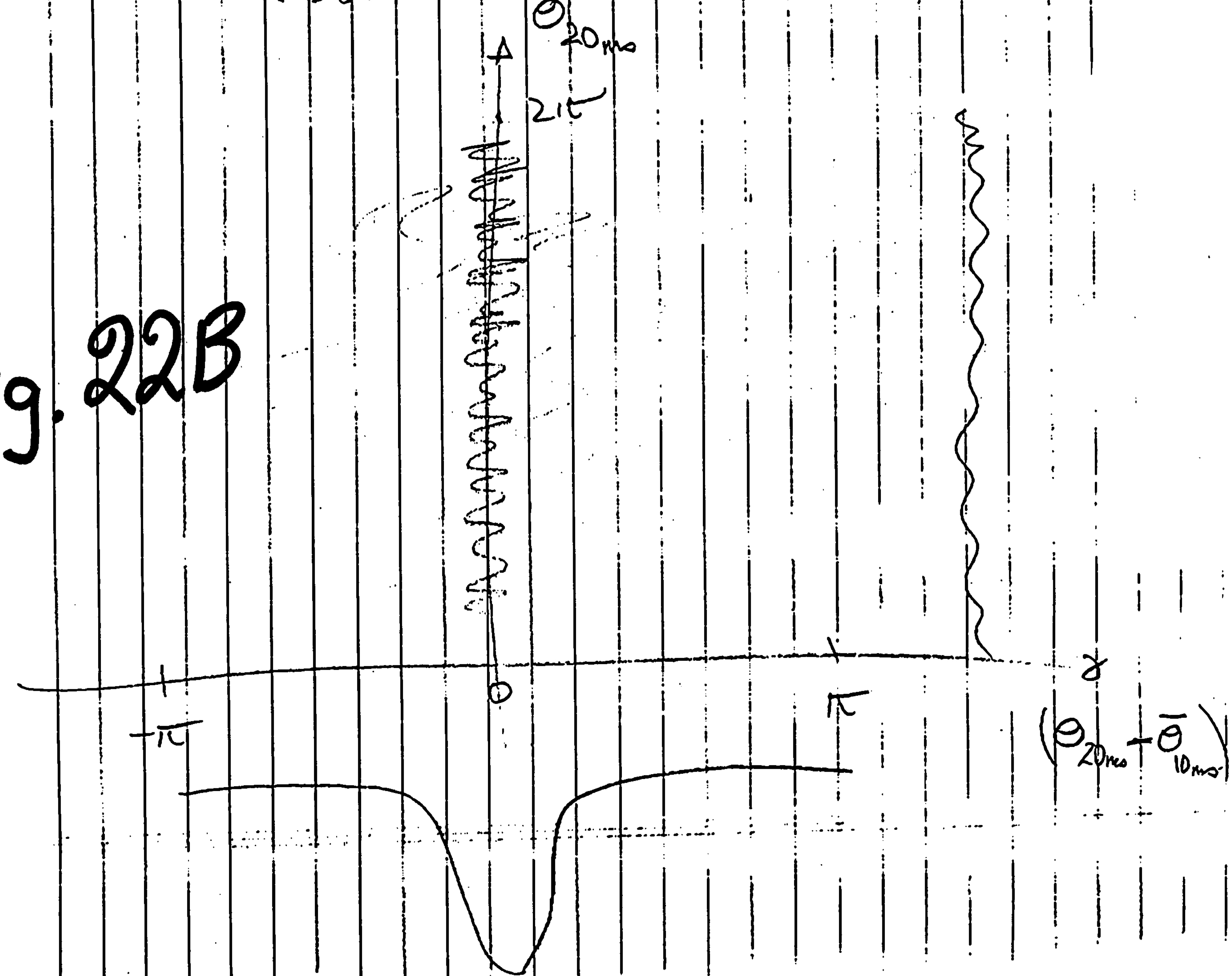
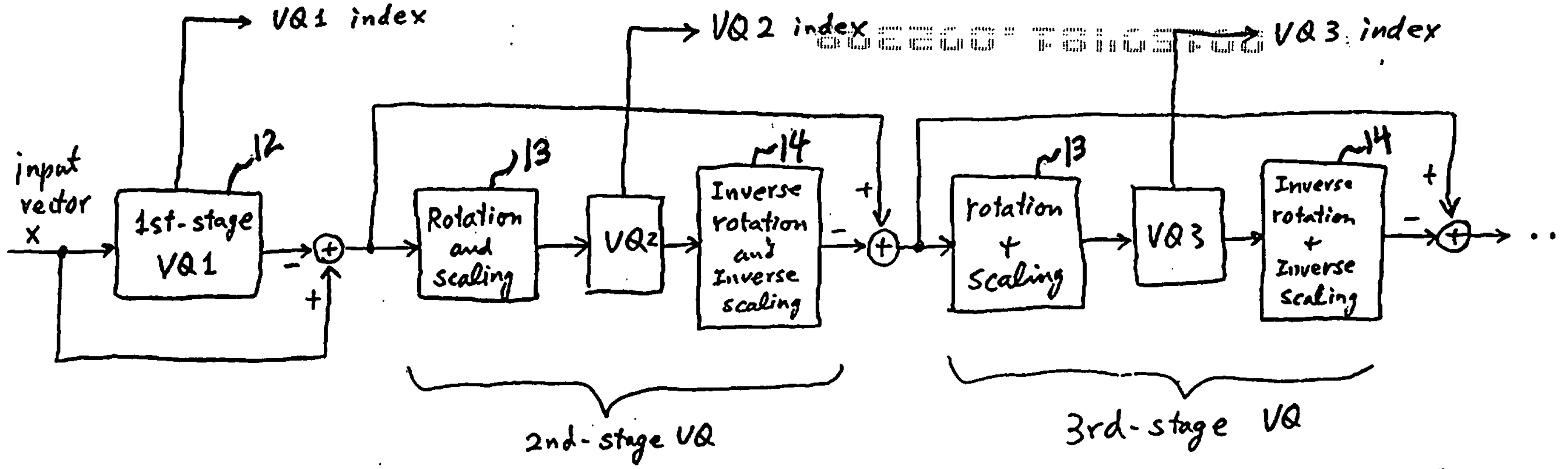
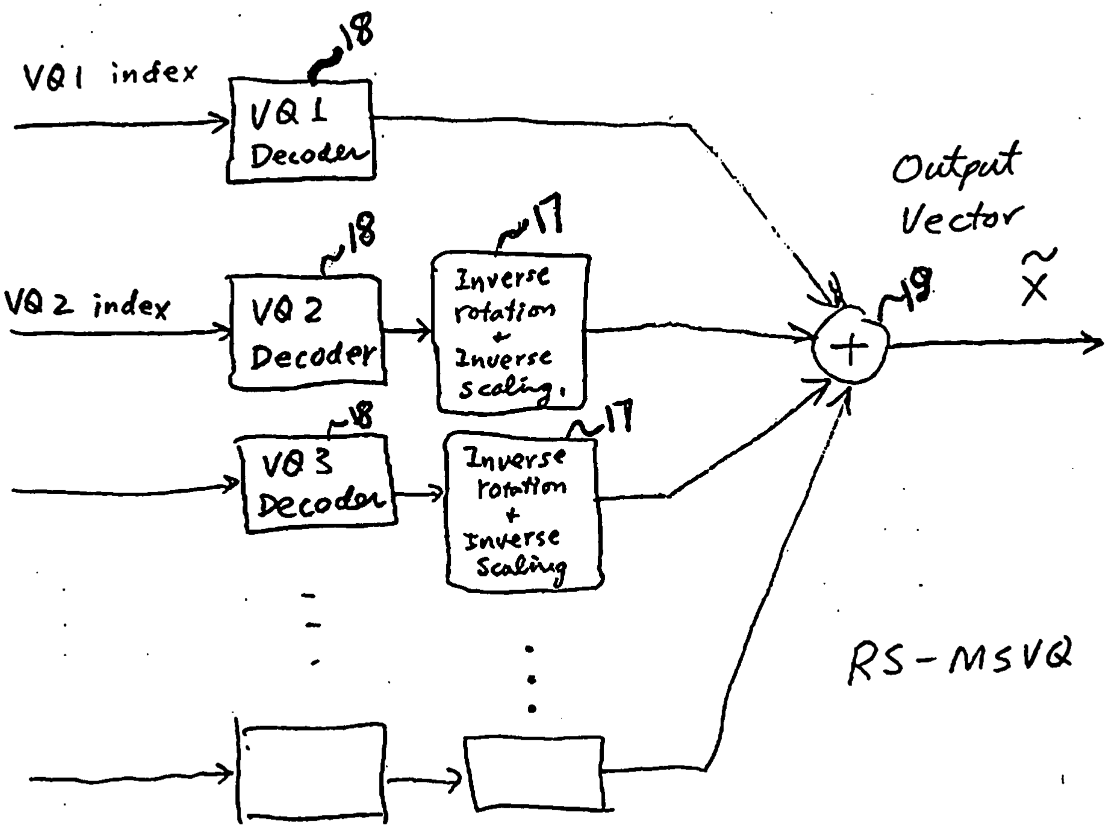


Fig. 22B



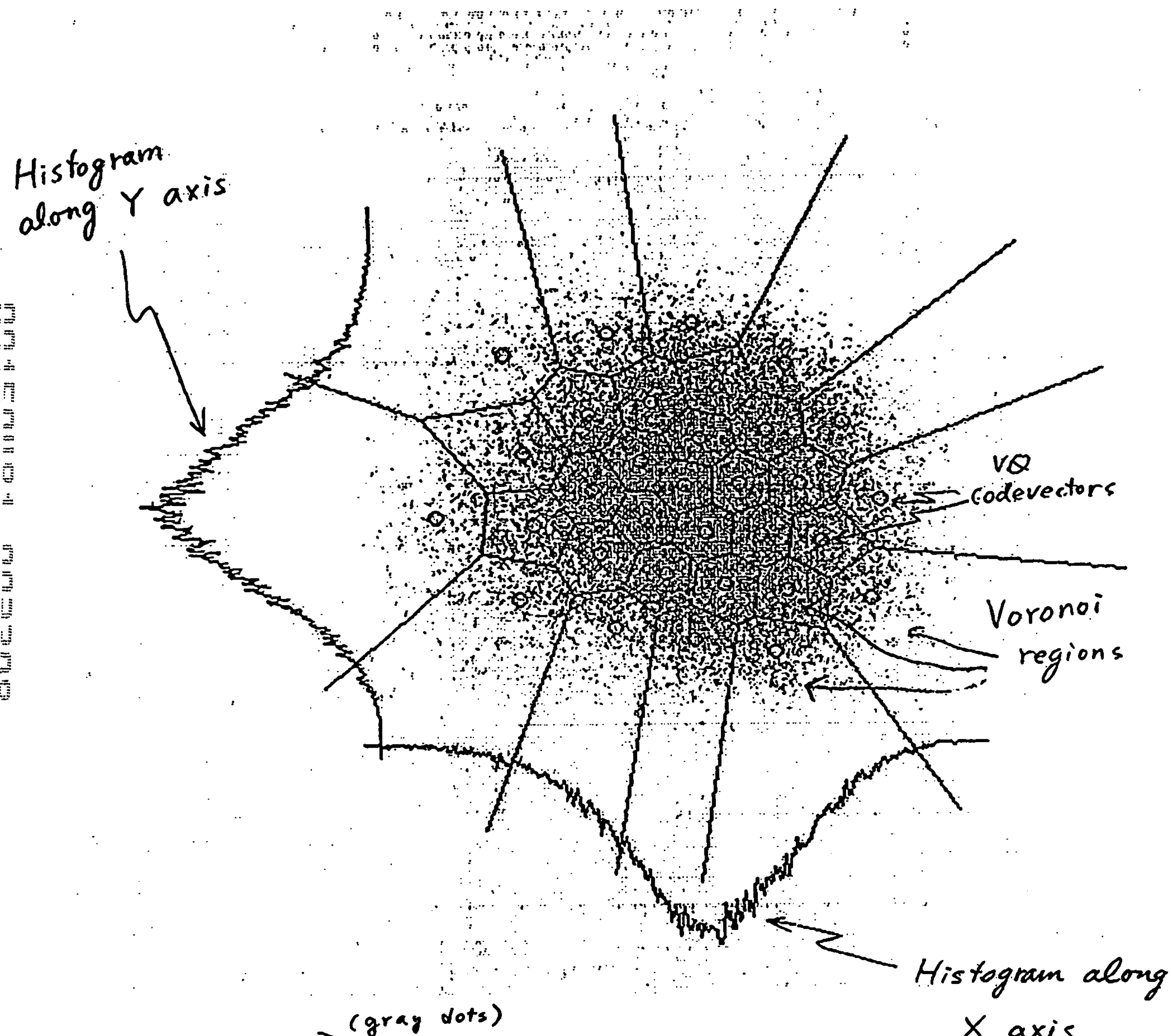
RS-MSVQ Encoder

Fig. 23A



RS-MSVQ Decoder

Fig. 23B



Scatter plot (gray dots) of 4th pair of ASIN(k) intra-frame prediction error, the histogram along each direction, and the corresponding 1st-stage 5-bit VQ codebook and Voronoi regions

Fig. 24A

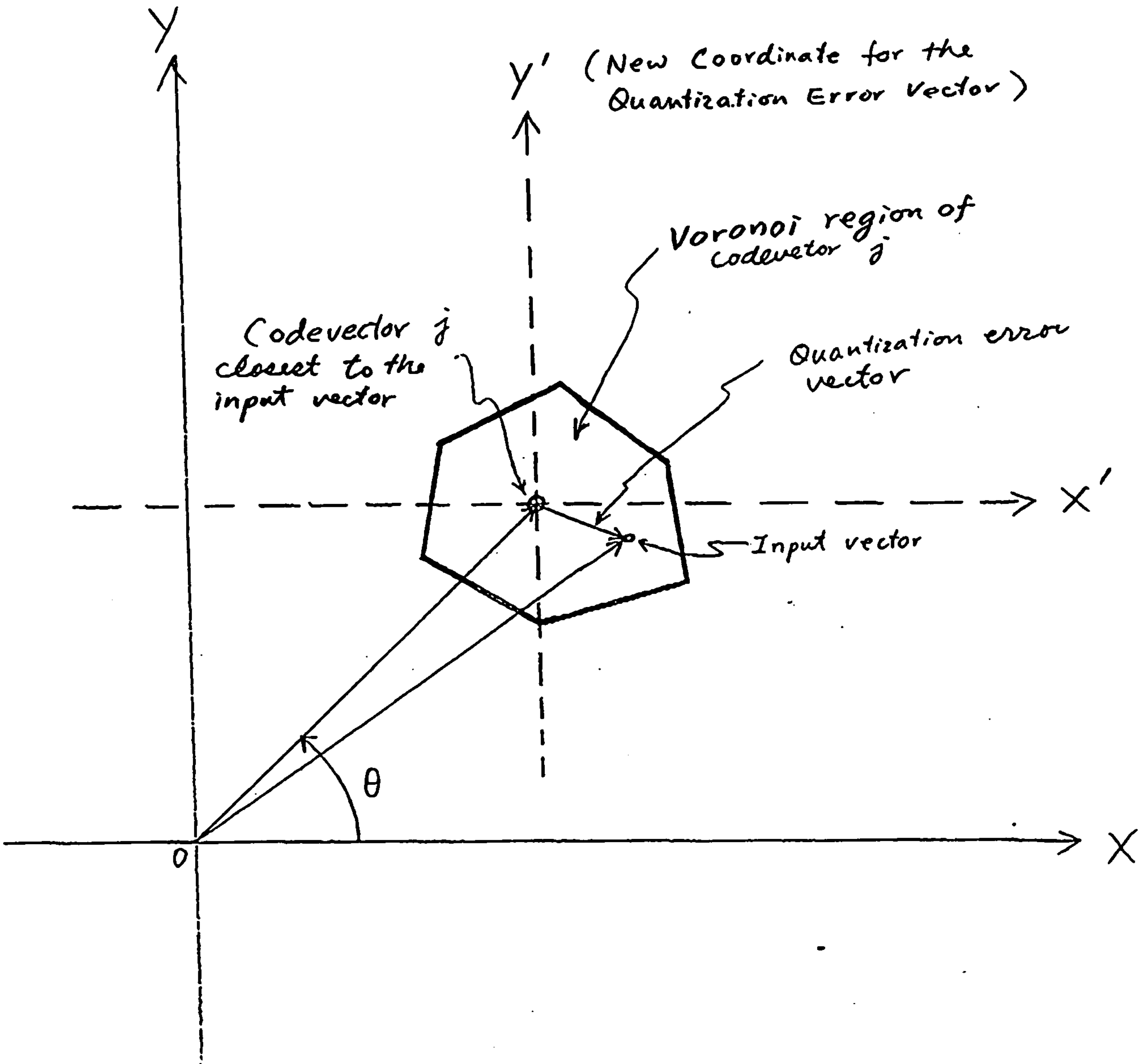
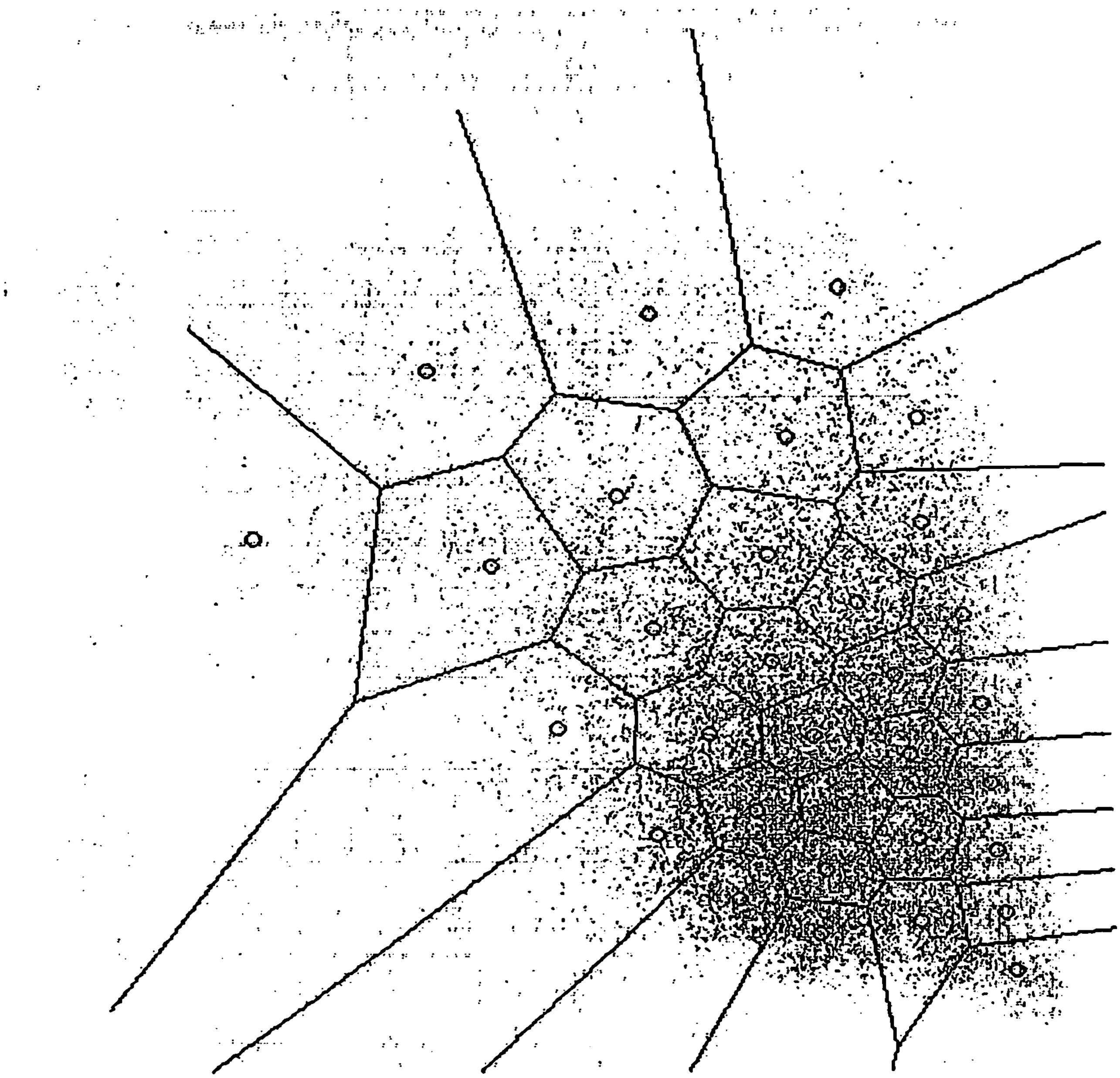


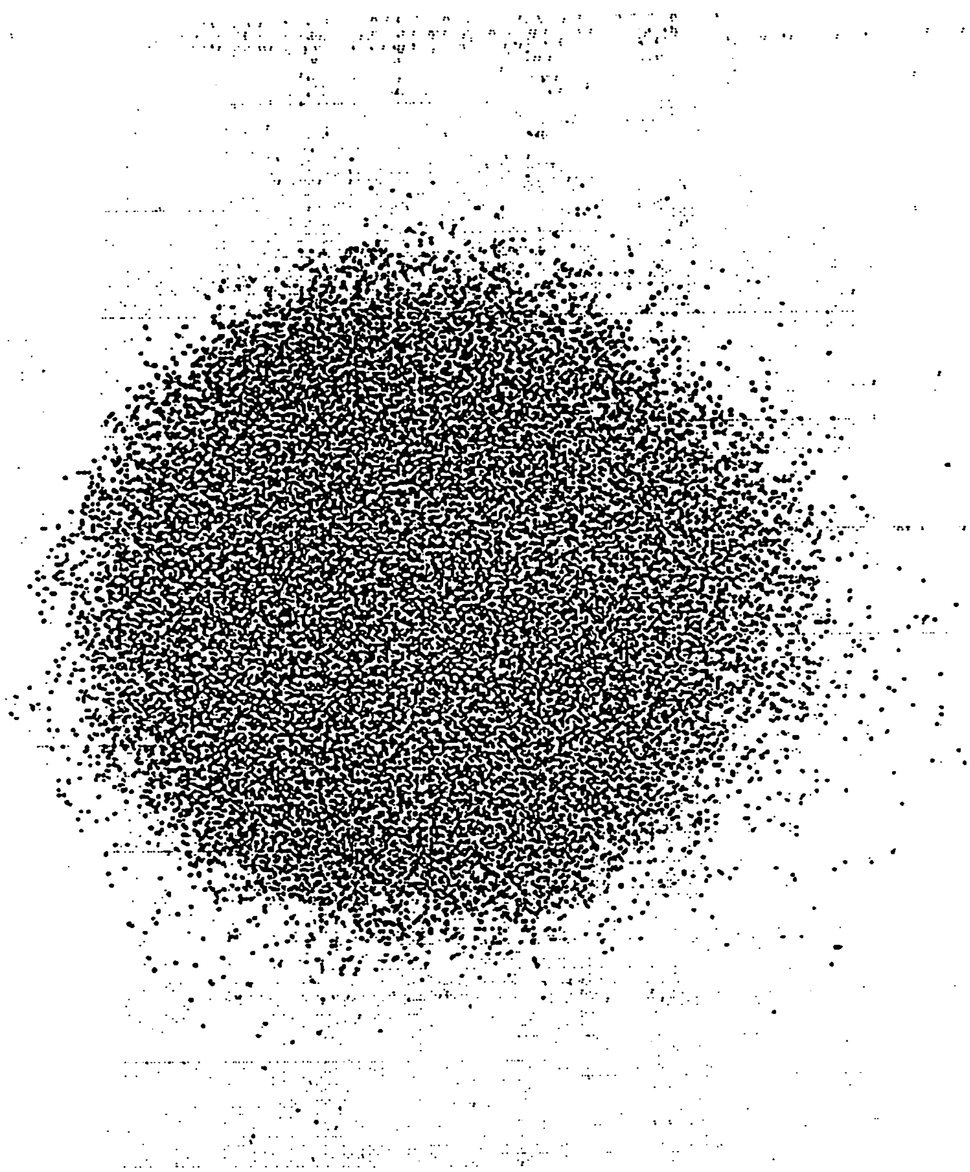
Fig. 24 B Illustration of the effect of subtracting the closest codevector for the input vector to get the quantization error vector.

Fig. 24B



Scatter plot of 1st pair of ASIN(k) (gray dots) and 1st-stage VQ codebook (small circles), and the corresponding voronoi cells

Fig. 24C

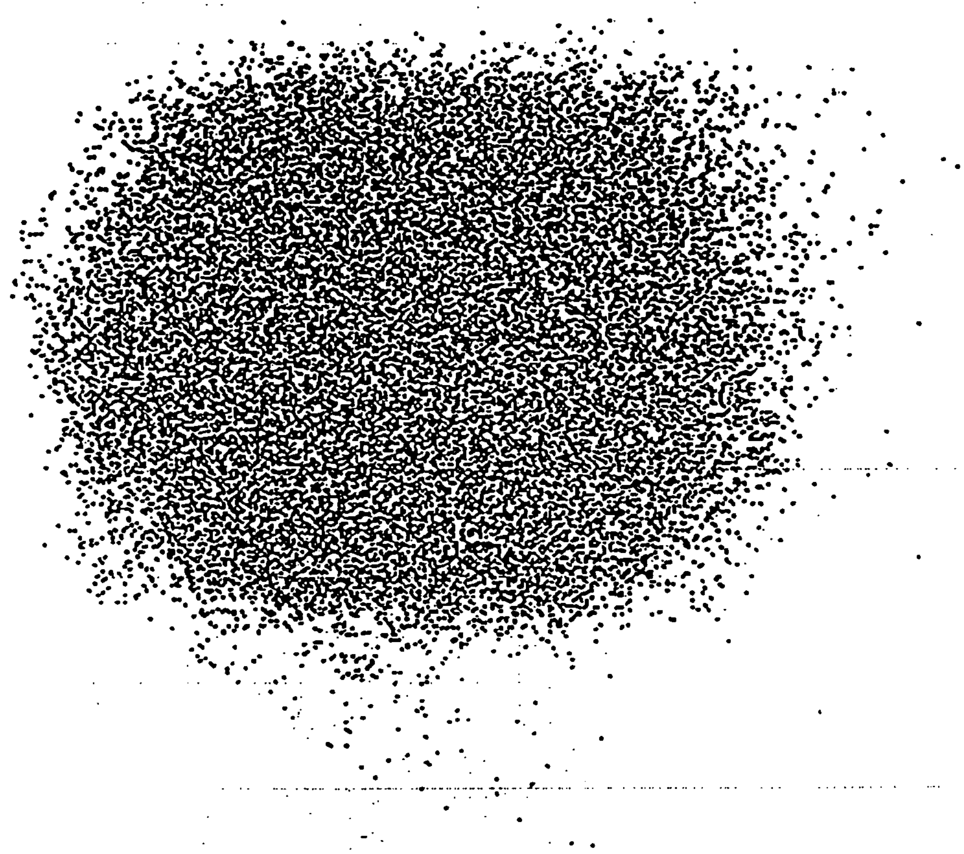


Without Hand-tuned Rotation Angles

—— inner cells of 1st pair of ASIN(k)

1st-stage VQ of

Fig. 25

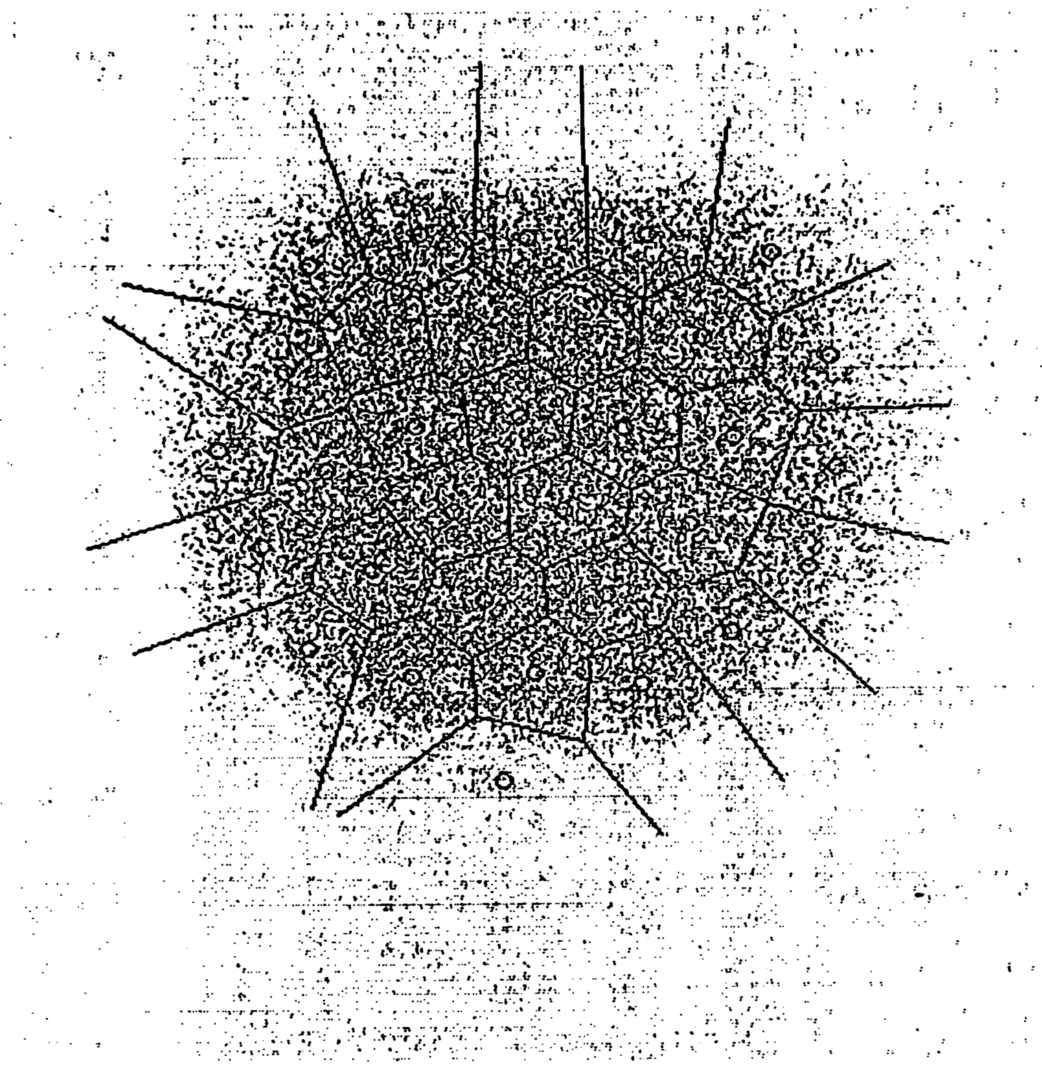


With hand-tuned rotation angles

— inner cells of 1st pare of ASIN(k)

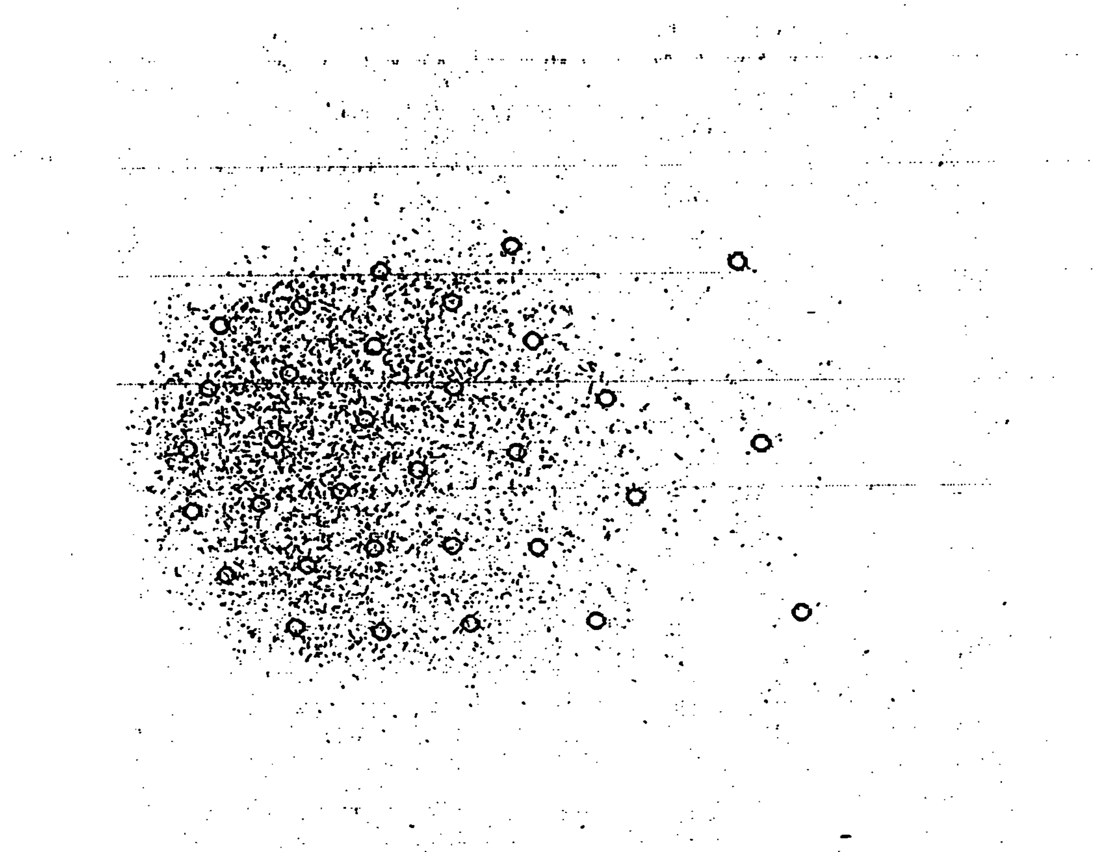
of 1st-stage VQ of

Fig. 26



Inner-cell 1st-stage VQ error vector distribution (gray dots)
(hand tuning) and corresponding 2nd-stage VQ codebook (small circles)
for 1st pair of ASIN (k)

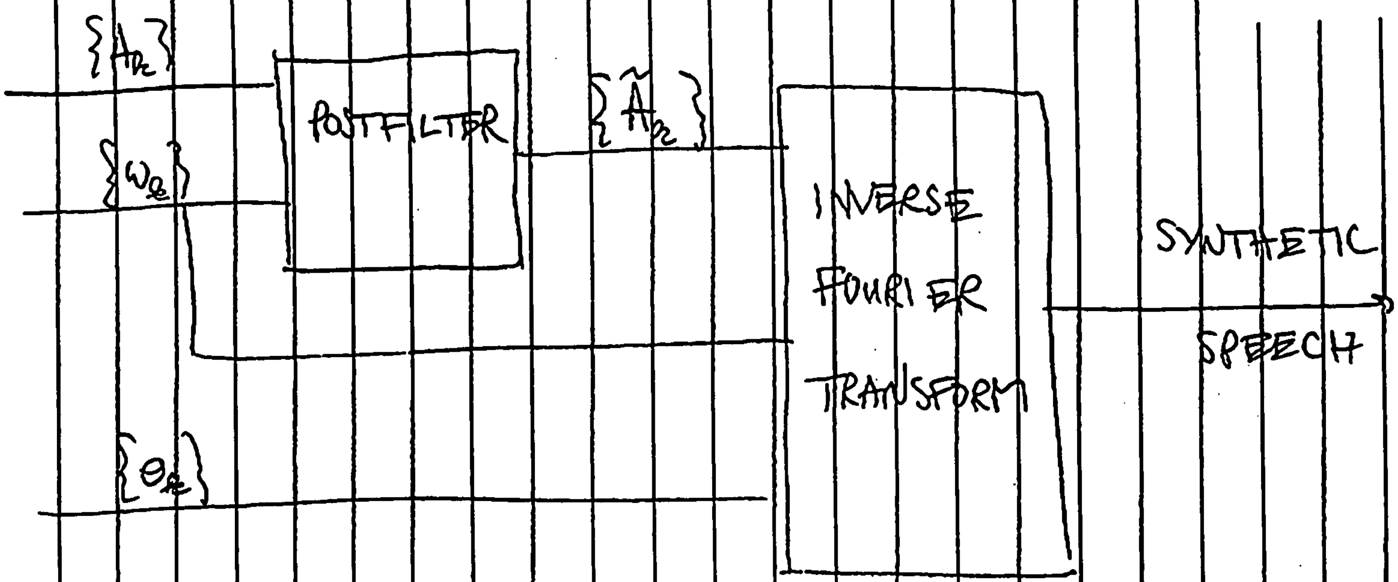
Fig. 27



Outer-cell 1st-stage VQ error vector distribution and corresponding 2nd-stage VQ codebook (small circles) for 1st pair of ASIN(k)

Fig. 28

SINUSOIDAL SPEECH SYNTHESIS



A_k = k^{th} sine-wave amplitude

ω_k = k^{th} sine-wave frequency

θ_k = k^{th} sine-wave phase

$\tilde{A}_k$ = k^{th} post-filtered sine-wave amplitude

Fig. 29

CANONICAL FREQUENCY-DOMAIN POSTFILTER

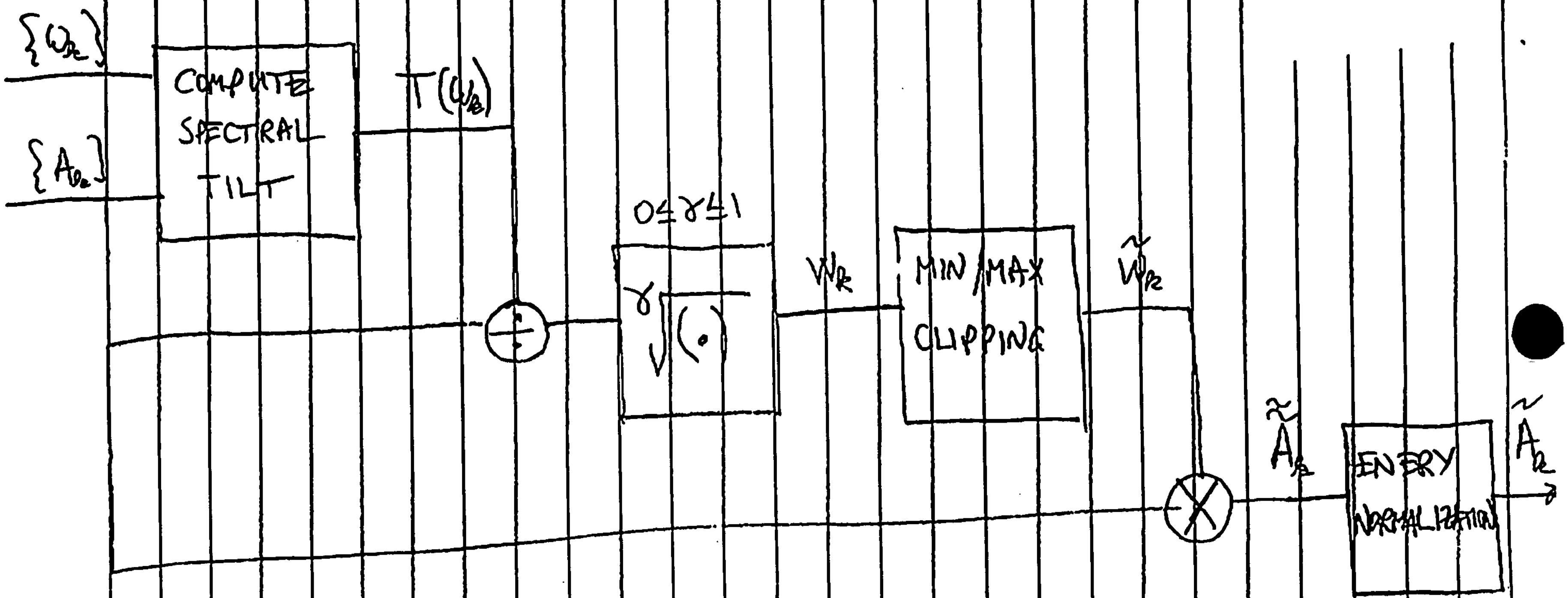


Fig. 30

CONSTANT COMPLEXITY POSTFILTER

$E(\omega)$ = amplitude envelope

ω_k = downsampled frequencies

$\tilde{E}(\omega)$ = post-filtered amplitude envelope

Fig. 31

CONSTANT COMPLEXITY POSTFILTER COMPUTED FROM CEPSTRAL COEFFICIENTS

C_m = m^{th} CEPSTRAL COEFFICIENT

$E(\omega)$ = AMPLITUDE ENVELOPE

$\tilde{E}(\omega)$ = POST-FILTERED AMPLITUDE ENVELOPE

Fig. 32

HAST CONSTANT COMPLEXITY POSTFILTER

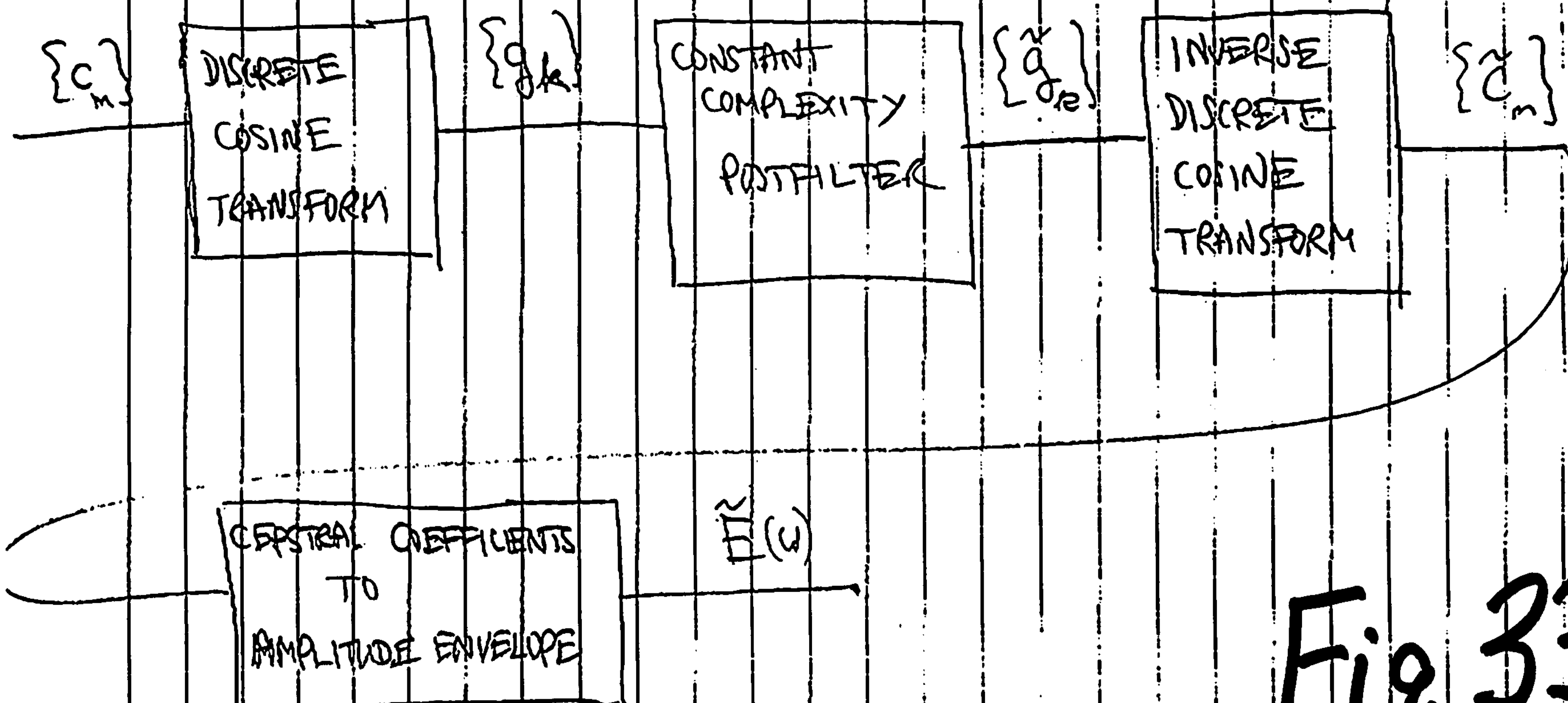


Fig. 33

c_m = m^{th} cepstral coefficient

g_k = k^{th} DCT coefficient = k^{th} channel gain

$\tilde{g}_k$ = k^{th} post-filtered channel gain

$\tilde{c}_m$ = m^{th} post-filtered cepstral coefficient

$\tilde{E}(\omega)$ = post-filtered amplitude envelope

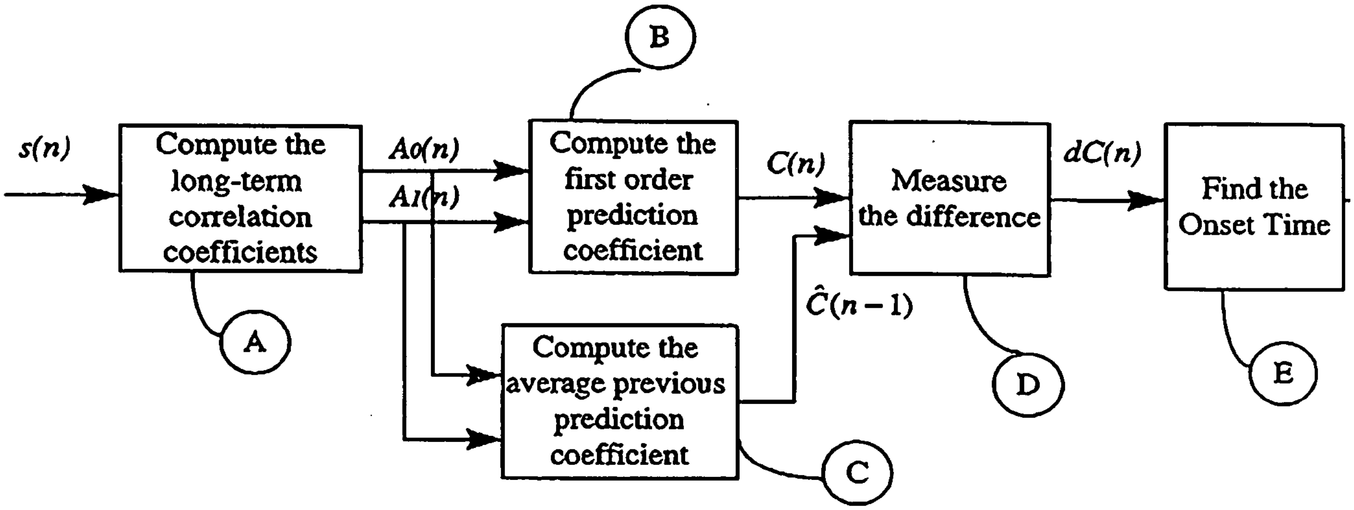


Figure 34.

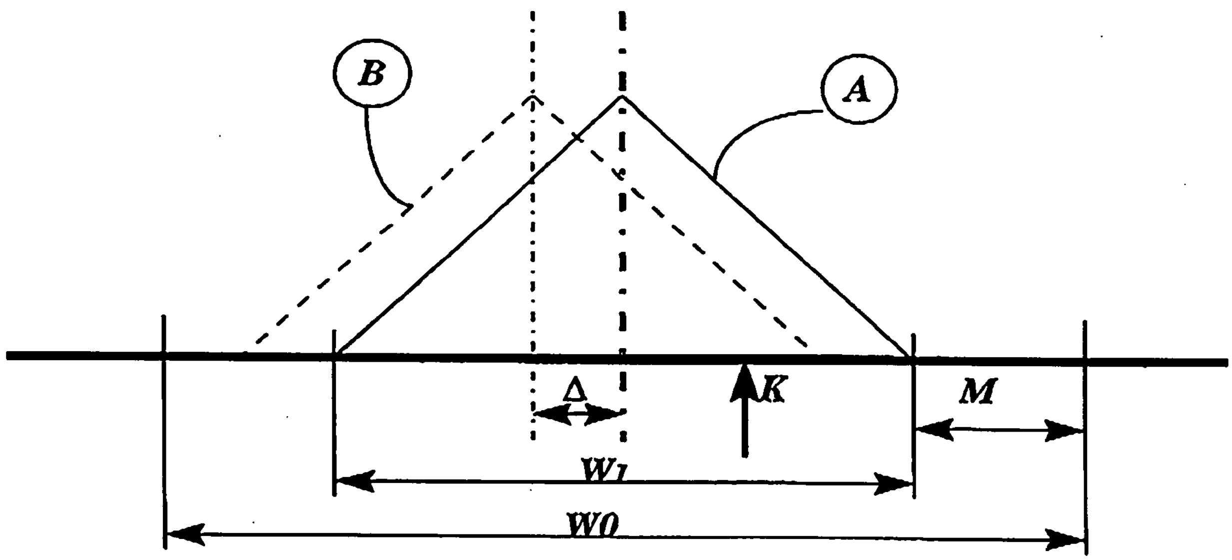


Figure 35.